

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2001, 16:36:28 ; Search time 3028.25 Seconds  
(without alignments)  
11538.356 Million cell updates/sec

Title: US-09-868-474-1

Perfect score: 2118  
Sequence: 1 ccgtgaagatgcgcctcc.....aaaaaaaaaaaaaaaaaaaaa 2118

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Geneml:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_ov:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pt:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_om:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_sy:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htgo\_hum:\*  
31: em\_htgo\_inv:\*  
32: em\_htgo\_rod:\*  
33: em\_htg\_hum:\*  
34: em\_htg\_inv:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

8

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2118	100.0	2118	9	AF195651 Homo sapi
2	2098	99.1	2143	9	AK000639 Homo sapi
3	1380.4	65.2	1399	9	BC000395 Homo sapi
4	1322.8	62.5	1637	9	AL050286 Homo sapi
5	1056.6	49.9	180268	2	AC023648 Homo sapi
6	1056.6	49.9	202333	2	AC087884 Homo sapi
7	868.8	41.0	202333	2	AF287293 Homo sapi
8	462	21.8	473	10	AF287293 Homo sapi
9	352	16.6	2335	9	AB055368 Macaca fa
10	258.4	12.2	274	6	AB055368 Macaca fa
11	197.8	9.3	180268	2	AC023648 Homo sapi
12	197.8	9.3	336508	2	AC011600 Homo sapi
13	152.8	7.2	336508	2	AC011600 Homo sapi
14	152.6	7.2	212282	2	AC023070 Mus muscu
15	60.4	2.9	2157	9	AK026630 Homo sapi
16	57.6	2.7	1813	8	AF229811 Cucumis s
17	57.4	2.7	2164	3	DDIAC715 Homo sapi
18	57.2	2.7	682	9	AK000414 Homo sapi
19	57	2.7	1649	8	ATW7450 Arabidops
20	56.2	2.7	1912	9	AB050421 Macaca fa
21	56	2.6	6108	12	ASPDAXHC Artificial
22	56	2.6	6134	12	ASPDAXHC Artificial
23	56	2.6	6146	12	AF269236 Cloning v
24	56	2.6	6872	12	AF269235 Cloning v
25	56	2.6	10452	12	AF269238 Cloning v
26	56	2.6	11166	12	AF269237 Cloning v
27	55.8	2.6	9934	3	DDU25660 Dictyostell
28	55.8	2.6	79594	9	AL138805 Human DNA
29	55.6	2.6	216410	2	AC027321 Homo sapi
30	55.4	2.6	3436	9	BC006201 Homo sapi
31	55.2	2.6	2488	9	AK027188 Homo sapi
32	55	2.6	384	6	AX185690 Sequence
33	55	2.6	1579	9	AF182416 Homo sapi
34	55	2.6	1896	5	AF024336 Dantio rer
35	55	2.6	3336	8	AF118843 Lycopersi
36	55	2.6	6105	12	ASPDAX3H Artificial
37	55	2.6	6113	12	ASPDAX3C Artificial
38	55	2.6	7218	6	I66494 Sequence 14
39	54.6	2.6	1316	9	AK027106 Homo sapi
40	54.6	2.6	3138	6	A21625 expression
41	54.4	2.6	754	9	HSR406933 Homo sapi
42	54.2	2.6	1953	9	AB056420 Macaca fa
43	54.2	2.6	227203	2	AC011184 Homo sapi
44	54	2.5	4898	3	DMU07629 Drosophila
45	53.6	2.5	1451	3	AF201906 Drosophila

## ALIGNMENTS

RESULT 1  
LOCUS AF195651 2118 bp mRNA PRI 14-APR-2001  
DEFINITION Homo sapiens cervical cancer 1 protooncogene protein p40 mRNA,  
complete cds.  
ACCESSION AF195651  
VERSION AF195651.1 GI:13624097  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 Kim,J.W.  
TITLE Identification of a new oncogene in human cancers  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 2118)  
TITLE Direct Submission  
JOURNAL Submitted (17-OCT-1999) Dept. of Obstet. & Gynecol., Kangnam St.





```

Db 998 GTGAAGATAGCTGGAACCTGGCTGGAGCAATGCGTGCAGATTTCTCGAGCTCAAG 1057
Qy 1021 aagctgagctgctctctctctgacaaagtgtccctgctccacaaactacattgga 1080
Db 1058 AAGCTGAGCTGTCTCTCTCTCTGACAACTGGTCTCTCTCTCTCTCTCTCTCTCTCT 1117
Qy 1081 caagcgcctgaatgaacatgagagcgatgcatgtccctgacgtatagatagacag 1140
Db 1118 CAAAGCCCTGAATGAACATGAGAGCGGATGTCCTCGACAGTGTATAGTATAGCAG 1177
Qy 1141 tgcagagaacaaacagcagctgcccagaaagtctgtgtatcgtttaagtgtggaagc 1200
Db 1178 TGCAGGACAAACAGCACTTGGCCAGAAAGTGTGTGTACTTAAAGTGTGGAGAGC 1237
Qy 1201 agagaagagcagagggccatgagctcacaagatgagcacccttgggaaatgagaca 1260
Db 1238 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1297
Qy 1261 ttcctctcaacatgaactgaacaaacccctctgtaggggtgtccgtgtgaggtgt 1320
Db 1298 TTCTCTCACAGCTAGACAGTGAACAAACCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1357
Qy 1321 caccctgcccctcaactaactaactaactaactaactaactaactaactaactaact 1380
Db 1358 CATCTGTCTCCCTCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1417
Qy 1381 tactgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1440
Db 1418 TACTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1477
Qy 1441 tatttgcttagagctgacccatccatcctcctcctcctcctcctcctcctcctcct 1500
Db 1478 TATTGTGCTTAGAGACTGATCCACTTCCATGTACTTACATCTGTGGGTTTGTGTGTGC 1537
Qy 1501 tcttgaacaaattttgtgctgtgtaaacacagcctccttctgtgctgagcactgtgtcat 1560
Db 1538 TCTTGAACAAATTTTGTGCTGTGTAACAAACAGCACTCTTGTGCTGTGAGCACTGTGCTGT 1597
Qy 1561 gaatgactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1620
Db 1598 GCAATGACTGTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1657
Qy 1621 ttttctgacactcaacgagcagtggtggaacactcttttcaacccctctgagacact 1680
Db 1658 TTTTCTGTACTCTTACAGTGGCTGATGAGCACTCTTCTTCAACCCCTCTGAAACCT 1717
Qy 1681 tttctgtaggtgaactgtgcaggaatattctatgtgaagataacagagagatcaagt 1740
Db 1718 TTTTCTGATGGGTACTTGCAGGAATATTCTTGAAGAAATGAACAGAAATGAAGTGAAGTGA 1777
Qy 1741 ctctctgagccctctcctcaatggtttctagccttcaactctcactgtcttctgtgagctgt 1800
Db 1778 CTCTCTTGAACCCCTTCTCTCAATGTTTCTTGAAGCTTCACTCTCATTTCTTCTGAGCTGT 1837
Qy 1801 attacagccctctgtgtatcttcaactctgtgctcactgagatgagcagcttcaact 1860
Db 1838 ATTACAGCCCTCTGTGGATCTTCAACTGTGTGCTCTCACTGTGATGATGATGATGATGAT 1897
Qy 1861 gtaactgacatgtgctgctctctgtggcactgatacaccctgttaagtaactaact 1920
Db 1898 GTAACCTGACATGTGCTGTCTCTTGTGGCCATGATGATGATGATGATGATGATGATGAT 1957
Qy 1921 gccagcctgtgagagatcaagagagagctctgatagttaagtaagttgggttttagctttct 1980
Db 1958 GCCCAGCCTGTGGGAGATCAGAGAGAGTCTGCATAGTATGATGATGATGATGATGATGATGAT 2017
Qy 1981 gctgtgactcagtgactagagctctgttaactatctgttaaatgataagagacgttt 2040
Db 2018 GTGTGATCACTAGTACTAGAGTCTGTATTAATTAATTAATTAATTAATTAATTAATTAATTA 2077
Qy 2041 ttaaccagtaagacactgtgtaaaactgtgtgataagaaaaaataaaaaaataaaaaa 2100

```

```

Db 2078 TTAACCCAAAGTAAAGACTGCTTGAACCTGTGATGAGAAAAAATAAAAAAATAAAAAA 2137
Qy 2101 aaaaaa 2106
Db 2138 AAAAAA 2143

RESULT 3
BC000395
LOCUS
DEFINITION
Homo sapiens, DKFZP586A011 protein, clone MGC:8483 IMAGE:2821851,
mRNA, complete cds.
ACCESSION
BC000395.1 GI:12653252
VERSION
MGC.
KEYWORDS
human.
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 1399)
REFERENCE
Strausberg, R.
Direct Submission
Submitted (15-NOV-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
MGC Project URL: http://mgc.nci.nih.gov
COMMENT
Contact: MGC help desk
Email: gcgabs-rt@mail.nih.gov
Tissue Procurement: DCTD/DTP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland.
Web site: http://www.nisc.nih.gov/
nisc.mgc@nih.gov
Contact:
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, O.L., Masello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stantrilop, S., Thomas, P.J.,
Tjongson, E.E., Touchman, J.W., Tsurgoun, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Series: IRAL Plate: 1 Row: 9 Column: 18
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4886472.

FEATURES
source
location/Qualifiers
1..1399
/organism="Homo sapiens"
/db_xref="LocusID:25875"
/db_xref="taxon:9606"
/clone="MGC:8483 IMAGE:2821851"
/tissue_type="lung, small cell carcinoma"
/clone_lib="NHL MGC 7"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
97..399
/codon_start=1
/product="DKFZP586A011 protein"
/protein_id="AAH00395.1"
/db_xref="GI:12653253"
/translation="MLTSTYLPPLLRRLKTRHTTVIHQDLAKALGIGOUTAOEYK
SACTLGLNSTHIGEDRCRTWGWENLQISLSEKAEISLLHNVLVLLSTNLTGTRR"

BASE COUNT
334 a 324 c 335 g 406 t

Query Match 65.2%; Score 1380.4; DB 9; Length 1399;
Best Local Similarity 99.6%; Pred. No. 0;

```

Matches 1384; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

QY 702 atcttgctctgaagagatgtttcttcaaccatctcttgagcatgaaccaactcagct 761
Db 10 ATCTTGCTCTGAAGAGATGTTTCTTAACCATCTCTGAGCATGAACCACTCCAGCT 69
QY 762 ttgcaatgaagaacttgagccgggacatgtcttcaacttcaactcctctcccttg 821
Db 70 TTGCACGTGAAGCCCTGAGCGGGCCATGCTTCAACATCTTACCTGCTCTCCCTTG 129
QY 822 ttgaaacatctgttgaagaatcatacaactgtgataccaactgagacaagcttgaca 881
Db 130 TTGAACATCTGTTGAAGCTCATCACTGATTTCACCACTGGAAGAGCTTTGGCA 189
QY 882 aagctggagatggagcagctgactctcaggaagtcaaaatcgagctgtatctcgtgac 941
Db 190 AAGCTGGGATGGCCAGCTGACTGCTCAGGAAGTAAATCGGCTTGTTATCTCCGTG 249
QY 942 ctgaatctcaagcataatgggaagaataaggtgtcgaacttgcttggaagaatgctgag 1001
Db 250 CTGAATCTTACCATATTTGTAAGATAGTGTGCAACTTGCTGGAGAAATGCTGCAG 309
QY 1002 attctctgagcctgaagaagctgagctgtctctcttgctgacacagtgctgctgc 1061
Db 310 ATTTCTCTGAGCTGAAGAAAGCTGAGCTGTCTCTCTGCTGCAACACTGTGCTCTC 369
QY 1062 tccaccaactactctgggagcaagcgctgaatgaaccatctgagagagctgtcctg 1121
Db 370 TCCACCAACTACCTTGGGCAAGCGGCTGAATGAACCTGAGCGGATGGCTTCTCG 429
QY 1122 cagctgataatagatagcagtgaggaagaacaacagcacttgcaagaagctgtgtgac 1181
Db 430 CAGCTGTATAGTATAGTACAGTCAAGACAAACAGCATTTGCCAGCAAGTGTGTGAC 489
QY 1182 ttttaagtgtgtggagagcagagaagaagaagcaggggcatatggcttcaacagatgaca 1241
Db 490 TTTTAAAGTGTGGAGAGGAGAGAGAGGCGAGGCGCATGGGCTTCAACAGATGGCACA 549
QY 1242 cctggagagcagcagacatctctctcagctagaactgaagaacaacccctcttgagg 1301
Db 550 CCTGGGAGACGACAGACTTCTCTCAGCTGAATGAATGAACAAACCTCTTCTTAG 609
QY 1302 ggtgagtcgtgaggtgtcactcctgccccctcaataactaaatagctggaacttgaca 1361
Db 610 GGTGTCCTGTGAGAGTGATCTGCTCCCTCAATAATTAATTAATTAATTAATTAATTA 669
QY 1362 gaagcctctacagcagcttactgtgagtgatgctcagctcagctgaagctgctt 1421
Db 670 GCAGCCTTACTGAGGCTTACTGATGTGTTCAGTTCATGCTCTAGGAATGTAGCTTT 729
QY 1422 tggcccaagtggaaatccttacttgctgaagcagatccacttccatgttacttacc 1481
Db 730 TGGCCAGGTGGGAAATCTTATTTGGCTTAGAGCATGATCCACTTCACTTACTTACTC 789
QY 1482 tgtgggtttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1541
Db 790 TGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 849
QY 1542 ctggagagcctgtgtcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatg 1601
Db 850 CTGGAGCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 909
QY 1602 caaaaatacaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1661
Db 910 CAAAATCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 969
QY 1662 tcaaacctctgaacaacttctctgataggtaactgtcaggaataatctatgtgaaag 1721
Db 970 TCAAAACCTCTGAACAACCTTTTCTGATGGGTAACTTGAGGAATATTTCTATTGGAAAG 1029
QY 1722 ataaacaggaagaagaagtgctctctgacccctctcctaagtgttcttcaactcactctcc 1781
Db 1030 ATAAACAGGAAGTAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1089

```

```

QY 1782 atgtctcttcctgggctgtatatacagccctctgtgatacttcaactcgtgctccact 1841
Db 1090 ATGTGCTTTTCTGGGCTGTATATACAGCCCTGTGGAATCTTCAACTCTGCTGCTCACT 1149
QY 1842 gtalgcagcagtcacacatgtaactgacagtgctgctctctctctgagccatgatacaca 1901
Db 1150 GTGATGCCAGCAGTCCAACTGTAACTGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1209
QY 1902 ctgtgaagtgataatctctgccccagcctggggagagatcaagaagagtgctgatagtga 1961
Db 1210 CTGTAAAGTAACTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1269
QY 1962 agttgggttgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2021
Db 1270 AGTTGGGTTTACCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1329
QY 2022 aatgcataagcactgtttttaaaccagtaaaagactgtctgaaactgtgtatgaaagaa 2081
Db 1330 AATGCATGAACACTGTGTTTAAACCAAGTAAAGACTGCTGAAACCTGAAACCAAAAAA 1389
QY 2082 aaaaaaaaaa 2091
Db 1390 AAAAAAAAAA 1399

```

#### RESULT 4

```

HSM800205 1637 bp mRNA PRI 18-FEB-2000
LOCUS Homo sapiens mRNA; cDNA DKFZP586A011 (from clone DKFZP586A011);
DEFINITION partial cds.
ACCESSION AL050286
VERSION AL050286.1 GI:4886472
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1637)
AUTHORS Koehler, K., Beyer, A., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
TITLE Direct Submission
JOURNAL Submitted (15-MAY-1999) MIPS, Am Kiofepspitz 18a, D-82152
Munich, Germany
COMMENT Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the Charité
Berlin/Germany) within the cDNA Sequencing consortium of the German
Genome Project.
This clone (DKFZP586A011) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.

```

#### FEATURES

source

```

1. 1637
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZP586A011"
/clone_1lb="586 (synonym: hute1). Vector pSport1; host
DH10B; sites NotI + SalI/MluI"
/dev_stage="adult"
/tissue_type="uterus"
331..633
/gene="DKFZP586A011"
331..633
/gene="DKFZP586A011"
331..633
/gene="DKFZP586A011"
/note="Similarity to D.melanogaster anon-66Db"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB43387.1"
/db_xref="GI:4886473"
/db_xref="SPTREMBL:O913X3"
/translation="MLTSLYLPPLLRHRLKTHITVHQDLAKLAKLIGQLTAEVK

```

```

gene
CDS

```

polyA_site	1620			
BASE COUNT	392 a	378 c	382 g	485 t
ORIGIN				

Smith, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/repeatmasker.html>  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: L6516  
 Center clone name: 624\_H\_4  
 ----- Summary Statistics  
 Sequencing vector: M13; M77815; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 15367 bases at least Q40  
 Consensus quality: 164187 bases at least Q30  
 Consensus quality: 170599 bases at least Q20  
 Insert size: 176000; agarose-fp  
 Insert size: 176868; sum-of-contigs  
 Quality coverage: 5.3 in Q20 bases; agarose-fp  
 Quality coverage: 5.3 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 35 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1 1101: contig of 1101 bp in length  
 \* 1102 1201: gap of 100 bp  
 \* 1102 1201: contig of 1057 bp in length  
 \* 2259 2358: gap of 100 bp  
 \* 2259 2358: contig of 1103 bp in length  
 \* 3462 3561: gap of 100 bp  
 \* 3462 3561: contig of 1078 bp in length  
 \* 3562 4639: gap of 100 bp  
 \* 3562 4639: contig of 1078 bp in length  
 \* 4640 4739: gap of 100 bp  
 \* 4640 4739: contig of 1691 bp in length  
 \* 4740 6430: gap of 100 bp  
 \* 4740 6430: contig of 1124 bp in length  
 \* 6431 6530: gap of 100 bp  
 \* 6431 6530: contig of 1124 bp in length  
 \* 6531 7654: gap of 100 bp  
 \* 6531 7654: contig of 1616 bp in length  
 \* 7655 7754: gap of 100 bp  
 \* 7655 7754: contig of 1616 bp in length  
 \* 9371 9470: gap of 100 bp  
 \* 9371 9470: contig of 1017 bp in length  
 \* 9471 10487: gap of 100 bp  
 \* 9471 10487: contig of 1017 bp in length  
 \* 10488 10587: gap of 100 bp  
 \* 10488 10587: contig of 1260 bp in length  
 \* 10588 11847: gap of 100 bp  
 \* 10588 11847: contig of 1579 bp in length  
 \* 11848 13526: gap of 100 bp  
 \* 11848 13526: contig of 1579 bp in length  
 \* 13527 13626: gap of 100 bp  
 \* 13527 13626: contig of 1599 bp in length  
 \* 13627 15225: gap of 100 bp  
 \* 13627 15225: contig of 1599 bp in length  
 \* 15226 15325: gap of 100 bp  
 \* 15226 15325: contig of 1694 bp in length  
 \* 15326 17019: gap of 100 bp  
 \* 15326 17019: contig of 1106 bp in length  
 \* 17120 18225: gap of 100 bp  
 \* 17120 18225: contig of 1106 bp in length  
 \* 18226 18325: gap of 100 bp  
 \* 18226 18325: contig of 2260 bp in length  
 \* 18326 20585: gap of 100 bp  
 \* 18326 20585: contig of 2260 bp in length  
 \* 20586 20685: gap of 100 bp  
 \* 20586 20685: contig of 1667 bp in length  
 \* 20686 22352: gap of 100 bp  
 \* 20686 22352: contig of 1667 bp in length  
 \* 22353 22452: gap of 100 bp  
 \* 22353 22452: contig of 3831 bp in length  
 \* 26284 26383: gap of 100 bp  
 \* 26284 26383: contig of 2178 bp in length  
 \* 28562 28661: gap of 100 bp  
 \* 28562 28661: contig of 2614 bp in length  
 \* 31276 31375: gap of 100 bp  
 \* 31276 31375: contig of 2614 bp in length  
 \* 31376 34050: gap of 100 bp  
 \* 31376 34050: contig of 2675 bp in length  
 \* 34051 34150: gap of 100 bp  
 \* 34051 34150: contig of 3225 bp in length  
 \* 34151 37375: gap of 100 bp  
 \* 34151 37375: contig of 3225 bp in length  
 \* 37376 37475: gap of 100 bp  
 \* 37376 37475: contig of 3230 bp in length  
 \* 40706 40805: gap of 100 bp  
 \* 40706 40805: contig of 3230 bp in length  
 \* 40806 44888: gap of 100 bp  
 \* 40806 44888: contig of 4083 bp in length

FEATURES  
 source  
 \* 44889 44988: gap of 100 bp  
 \* 44989 51688: contig of 6700 bp in length  
 \* 51689 51788: gap of 100 bp  
 \* 51789 60033: contig of 8245 bp in length  
 \* 60034 60133: gap of 100 bp  
 \* 60134 66633: contig of 6500 bp in length  
 \* 66634 66733: gap of 100 bp  
 \* 66734 72372: contig of 5639 bp in length  
 \* 72373 72472: gap of 100 bp  
 \* 72473 80057: contig of 7585 bp in length  
 \* 80058 80157: gap of 100 bp  
 \* 80158 89044: contig of 8887 bp in length  
 \* 89045 89144: gap of 100 bp  
 \* 89145 99209: contig of 10065 bp in length  
 \* 99210 99309: gap of 100 bp  
 \* 99310 109103: contig of 9800 bp in length  
 \* 109110 109209: gap of 100 bp  
 \* 109210 118651: contig of 9442 bp in length  
 \* 118652 118751: gap of 100 bp  
 \* 118752 130023: contig of 11272 bp in length  
 \* 130024 130123: gap of 100 bp  
 \* 130124 140909: contig of 10786 bp in length  
 \* 140910 141009: gap of 100 bp  
 \* 141010 154860: contig of 13851 bp in length  
 \* 154861 154960: gap of 100 bp  
 \* 154961 180268: contig of 25308 bp in length.  
 Location/Qualifiers  
 1..180268  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="12"  
 /map="12"  
 /clone.lib="RPCT-11 Human Male BAC"  
 1..1101  
 /note="assembly-fragment"  
 1202..2258  
 /note="assembly-fragment"  
 2359..3461  
 /note="assembly-fragment"  
 3562..4639  
 /note="assembly-fragment"  
 4740..6430  
 /note="assembly-fragment"  
 6531..7654  
 /note="assembly-fragment"  
 7755..9370  
 /note="assembly-fragment"  
 9471..10487  
 /note="assembly-fragment"  
 10588..11847  
 /note="assembly-fragment"  
 11848..13526  
 /note="assembly-fragment"  
 13527..15225  
 /note="assembly-fragment"  
 15226..17019  
 /note="assembly-fragment"  
 17020..18225  
 /note="assembly-fragment"  
 18226..20585  
 /note="assembly-fragment"  
 20586..22352  
 /note="assembly-fragment"  
 22353..26283  
 /note="assembly-fragment"  
 26284..28561  
 /note="assembly-fragment"  
 28562..31275  
 /note="assembly-fragment"  
 31276..34050  
 /note="assembly-fragment"  
 34051..37375  
 /note="assembly-fragment"  
 37376..40705  
 /note="assembly-fragment"  
 40706..44888  
 /note="assembly-fragment"





```

OY 1979 ggtgtgcatcagtgacttaagatctgtgtaacttaattgttaaatcatgaagactgt 2038
|||||
Db 70211 GTGTGTGATGATGACTTATGAGTTCTGTATTAATCTATTTGTAATGATGATGACACTGT 70270
OY 2039 tttaaacccaagttaagactgcttgaacctgttgaatgagaa 2081
|||||
Db 70271 TTTTAAACCAAGTAAAGACTGCTGTAACCTGTTGATGAGAA 70313

RESULT 7
LOCUS AF287293 2447 bp mRNA ROD 02-AUG-2001
DEFINITION Mus musculus cervical cancer receptor (Mccr) mRNA, complete cds.
ACCESSION AF287293
VERSION AF287293.1 GI:15077021
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Hwang, J.H., Kim, J.W. and Hwang, S.Y.
TITLE Mouse cervical cancer receptor
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2447)
AUTHORS Hwang, J.H., Kim, J.W. and Hwang, S.Y.
TITLE Direct Submission
JOURNAL Submitted (13-JUL-2000) Biochemistry, Hanyang University, 1271
Sa-dong, Ansan, Kyunggi-do 425-791, Korea
FEATURES
Source
1..2447
/organism="Mus musculus"
/strain="C3H/He"
/db_xref="taxon:10090"
/cell_line="p19"
/cell_type="embryonic carcinoma"
1..2447
/gene="Mccr"
31..1113
/gene="Mccr"
/codon_start=1
/product="cervical cancer receptor"
/protein_id="AAK83032.1"
/db_xref="GI:15077022"
/translational="MALSRCVMAARAALMGSTVAPGPFVTRQLDGRSGPAMRAPRSSK
LHLSPRKADVNILSYVYTKRAINGSYHREHLRQFRODTKCLFVLLISPPRANIVLLMYL
ARRKLAAMKRONLKHOLSYREMEHLRQFRODTKCLFVLLISPPRANIVLLMYL
FPKRLVKNHWTPEQJIDFLDYVHGLRSHSEVITHLRASFTVSHKRLRDLCL
TKVOSGTHPAAOVLAIRDCEFTYPLGFSQLOSKMALSOAMLLTPYLPPLRLRL
KSHYTHOLDRAKLAKLIGQLTAOEYKSNACYLRLGNSTHIADDRCLRAMLGEMLIHISC
SLKEPELSTLLHNVLLSTNYLETRR"
BASE COUNT 535 a 683 c 599 g 630 t
ORIGIN

```

Query Match 41.0%; Score 868.8; DB 10; Length 2447;  
 Best Local Similarity 81.6%; Pred. No. 1,9e-221;  
 Matches 1057; Conservative 0; Mismatches 227; Indels 12; Gaps 4;

```

Db 203 CTTAGCTGTGACCAAGACAGACGATTAACGGATCGTACCATCTTTCTTGGGCGCTC 262
OY 241 atttccccgcctctatccctgttacacaatcttaagaagatgagatgataag 300
Db 263 ATTTCCCGCCTCTCTATGCTTTATACAACTTCTATGAAGAAATACAGATGATGG 322
OY 301 ctgattccaaaagagctagaagaataaagaacaatatgtgagagcaaalataaagtttc 360
Db 323 CTGATGGCAAAAAGCGTAGAAGATTAAGGACAGATGTGAAACAAACAAACGTAAGTTTC 382
OY 361 atcaattccataccgggagatgagagatlttgagacagttccgccagaagctcacaagt 420
Db 383 ATCAACTCTCTTACCGGGAGATGAGACATTTGAGACAGTTCCGACGAGATATCACCAAT 442
OY 421 gttcttccaggtatattatccatccacctttgcaacacacagctggtctctgtcga 480
Db 443 GTCTCTTCCGTGGGCTCATCTCCATTCACACCTTTCGCAACTACCTGGCTTTTGTCTAA 502
OY 481 tgaactgtttcccaagcaactatgatcagagcttctggaaccccaacaacaacaactg 540
Db 503 TGATACGTGTTCTTAGGCAACTGCTAGTCAGATTTTGACACCCCAACAACAATTTG 562
OY 541 attcttagatatactatcatgtcttccggaagcagtcaccaagaataattagttatt 600
Db 563 ATTTCTTGACGTTTATCATGGTCTCCGAGGGGCTCCACTCGGAAGTCATCACACATC 622
OY 601 tagaagaagtcacccctccatcattctgagtcagagactcgggagcgtgagacagtcgt 660
Db 623 TAGAAGAGGCGCACACTTTTGTGTTCTCATGACAAATCCGAGGAGCTGACAGACTGT 682
OY 661 gcaacaagatatacagctggtglaaccacccagcaaatatatactgtgctctggaagat 720
Db 683 GCACCAAGGTGACAGATGGGACACCACCCAGACAGAGACGACTCGCTGCAGAGATT 742
OY 721 gttctctaacctcctctcgtggagcatgaaccaactcagagctltgcaagtgaagcttga 780
Db 743 GCTTCTTACCTACCTCCCTTGGGGCTTCAGCCACTCCAACTCCAGATGAGAGACTTGA 802
OY 781 gccggccaatgctctcaatcttaactcctcctcctcctctgttgaagacatggtgaaga 840
Db 803 GCCAGGCGATGCTCTCACACACTTACCTGCCCCCGCTTGGTGAAGGACGGGCTAAGA 862
OY 841 ctcaatacaactgtatcacaactcgaactggaacagctlttgcaaaagcttggagattggcagc 900
Db 863 GTACACACACCGGATCCACACACACTGAGACAGGCTCTGGCAAAAGCTGGGATTTGGCCAC 922
OY 901 tgcgtctcaggaagaataatgagcttgttatctccgttgctgtaatttaacgatatgt 960
Db 923 TGACTGCACAGGAAGTGAATCGGCTTGCTATCTTGTTGGTGAACCTCTACCCACATTG 982
OY 961 gtaagaataagtgctgaaactgtgctggaagatgtgtcagaatttccctcagcctgaag 1020
Db 983 CTGACGACAGGTCTCGAGCTGCTGGGGGAATGTTGTCACATCTCTTACCTGTAAG 1042
OY 1021 aagctgagctgtctctctgtgtgacaaagctgtgctctcacaacaaacttaacttggga 1080
Db 1043 AACCCGAGCTGCTCTCTGTGCTGACACAGGTGCTGCTGCTGCTGCAACACTACCTTGA 1102
OY 1081 caagcgctggaatgaa--ccaatgagagatgagatgtctcgcagctgataatagc 1138
Db 1103 CAAGGCGCTGAGTGAACCCCGAAGAGAGACGACATTTCTCGACGTGACCTAGTGTAGT 1162
OY 1139 agtcgaggaacaaacagacacttgcagcaagatctgtgtgtaactgttaagtgtgagag 1198
Db 1163 GGTGTGGACCAAGAACACTGCTCAGC--TGCTGTGTGGCACCACTA-----GTGTG 1213
OY 1199 gcaagagaagacagagggccatgggcttcaagcaatggaacagcaactgttggagatgaca 1258
Db 1214 GAGAGCAGCGAGATGGTTTCCACAGGGCTTCAAGACGCTATACCCCTCGGAGCTATAGA 1273
OY 1259 catctctcacaagctagaactgaaacaaacctct 1294
Db 1274 CATTCCT-TCACAGCTTGAGCCTAACAACCATGT 1308

```

RESULT	8
HST000016	
LOCUS	HST000016      473 bp    mRNA                  PRI         15-JUN-1999
DEFINITION	Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 138851.
ACCESSION	AL079293
VERSION	AL079293.1    GI:5102750
KEYWORDS	FLI CDNA.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 473)
REFERENCE	AufRAY,C., Ansoerge,W., Ballabio,A., Estivill,X., Gibson,K., Lehrach,H., Poustka,A. and Lundeberg,J. The European IMAGE consortium for integrated Molecular analysis of human gene transcripts Unpublished 2 (bases 1 to 473) Baseli,M.T., Banfi,S., Riboni,M., Ballabio,A. and Borsani,G. Direct Submission Submitted (14-JUN-1999) Telethon Institute of Genetics and Medicine (TIGEM), Via Olgettina 58, 20132 Milano, ITALY. Tel: ++39-02-215601 Fax: ++39-02-21560220 WWW site: http://www.tigem.it e-mail enquiries: basel@tigem.it, borsan@tigem.it EURO-IMAGE Consortium Contact: AufRay C
TITLE	JOURNAL
REFERENCE	
AUTHORS	
TTLE	
JOURNAL	
COMMENT	

This clone is available royalty-free through IMAGE Consortium Distributors.

**IMPORTANT:** This sequence represents the full insert of this IMAGE cDNA clone. No attempt has been made to verify whether this corresponds to the full-length of the original mRNA from which it was derived.

FEATURES	source
Location/Qualifiers	1. .473
/organism="Homo sapiens"	/db.xref="taxon:9606"
/chromosome="12"	/clone="IMAGE CDNA clone 138851"
/clone_1lb="Soares placenta (Nb2HP)"	/map="D12S333-D12S325"
BASE COUNT	151 a 94 c 92 g 136 t
ORIGIN	

Query Match	21.8%	Score 462	DB 9	Length 473
Best Local Similarity	100.0%	Pred. No.	1.1e-112	
Matches 462; Conservative	0	Mismatches	0	Gaps 0

OY	1657	cttttcaaaccccttaaacaccttttctgaagggaactgcgaagaatctcatgtg	1716
Db	2	CTTTTCAAAACCTCTGAACACCTTTTTCGATGGGTACTTTCAGGAATATTCATTTG	61
OY	1717	aaagataaacaggaagatcaagaagtctcttgaaccccttctcaatagtcttagccttacc	1776
Db	62	AAAGATATACAGGAGATACAAAGTCTTCTTGACCCCTTCTCAATGTTTATGCTTAC	121
OY	1777	tctccatgtctcttctcggagctgtatatacagccctctgtagtctcttcaactcgtgcct	1836
Db	122	TCTCCATGTCCTTTCTCGGCTGTATTACAGCCCTCTGAGATCTTCAACTCTGTGCTT	181
OY	1837	ccacgtgatacagcagttccaactcgaactgacaagcgtgctcctctctcggccatgata	1896
Db	182	CCACGTGATGCACACATTCACAGTATACAGAGTGGCTTGCTTCTCTGGGCCATGGAT	241
OY	1897	caacactgttaagtactaatctaactgcccagcctcggagagataagaagagctctgcatagt	1956

D <sub>b</sub>	242	CACACCTGTAGTGACTAATTACTGACCACCTGGGGAGATCAGAGAGSCTGCATAST	301
Q <sub>y</sub>	1957	tagtaagttgagtgtagcttltgtgtgtgcatacgaigacttagagtctgtataacta	2015
D <sub>b</sub>	302	TAGTAAGTTGGSGTTTACTCTTTGTGTGCATGATGACTTAAGATTCGTGATAACTTA	361
Q <sub>y</sub>	2017	tftgaaatagcgtgaacagctttttaaccagaatgaacagtcgtgaacctgtgat	2076
D <sub>b</sub>	362	TTGTAAATGCGATGAAGCATGCTGTTTTTAACCCAGTAAGTAAACGCTGTGAACCTGTGAT	421
Q <sub>y</sub>	2077	ggaaaaaagaaaaaagaaaaaagaaaaaagaaaaaagaaaaa	2118
D <sub>b</sub>	422	GCAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA	463

RESULT	9
AB055368	
LOCUS	AB055368 2335 bp mRNA PRI 07-FEB-2001
DEFINITION	Maaca fascicularis brain cDNA, clone:Of1A-12629.
ACCESSION	AB055368
VERSION	AB055368.1 GI:12698272
KEYWORDS	fis (full insert sequence).
SOURCE	Maaca fascicularis adult male frontal lobe left cDNA, to mRNA,

ORGANISM  
Macaca fascicularis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi  
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
Cercopithecinae; Macaca.  
1 (sites)

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Unpublished  
2 (bases 1 to 2335)  
Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.  
Direct Submission  
Submitted (02-FEB-2001) Katsuyuki Hashimoto, National Institute of  
Infectious Diseases, Division of Genetic Resources; 22-1, Toyama  
1-chome, Shinjuku-ku, Tokyo 162-8640, Japan  
(E-mail: khashi@nih.go.jp, URL: <http://www.nih.go.jp/jp/yoken/genebank/>  
Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)  
Lab host: TOP10

**R Site2:** DraIII (CACCAGTG)  
Description: 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCTTTTTTTTTTTTT]; double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pml185-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing (5' end primer [CTTGTGCTCTAAAGCTGCG]; 3' end primer [GCAGCTGAGGCTCGAGACAA]).

```

source
1..2335
/organism="Macaca fascicularis"
/db_xref="taxon:9541"
/clone="QfIA-12629"
/sex="male"
/tissue_type="frontal lobe left"
/clone_lib="macaque brain cDNA library QfIA"
/dev_stage="adult"

BASE COUNT      718 a      459 c      480 g      678 t

ORIGIN

Query Match      16.6%  Score 352;  DB 9;   Length 2335;
Best Local Similarity 95.2%  Pred No. 4e-83;
Matches 396;  Conservative 0;  Mismatches 15;  Indels 5;  Gaps 3;
```

```

Qy 1707 attcattggaagaatacaccagaagtaacagtgctctcttgacccctctcctcaatgtctc 1766
      |||||||
Db 1863 ATTCTATGGAAGAATAACAGAGTAACAGTGTCTTCTTACCCCTCTCTCAATGTCTTC 1922
Qy 1767 tagccttaacctcctcatgtctctctctctctctctctctctctctctctctctctca-- 1824
      |||||||
Db 1923 TAGCCTTACCTCTCTCATGTCTTCTTCCGCGCTGTATTACA-CCCCCTGTGATCTTACG 1981
Qy 1825 actctgctgctccacactgtatgtacagcagctgaactgaactgaactgaactgtctc 1884
      |||||||
Db 1982 GCTTTGGCTGCTCCACCTGTATGACAGCAATCACTGTAAGTGAAGTGGCTGCTTCTT 2041
Qy 1885 tgggccaatgatacacaactgtatgaactaactctgcccagcctgggagatcaaggaga 1944
      |||||||
Db 2042 TGGCCATGATGATCAACACTGTAAGTACTAATTAATGCCCCAGCTGGGAGATCAGAGAGA 2101
Qy 1945 ggtctgcat--agttagaaagtggtgttgctctctctctctctctctctctctctctc 2002
      |||||||
Db 2102 GGTCTGCAATCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2161
Qy 2003 tctgtaataactattgtatgaatgcatgaagcagctctctctctctctctctctctctctc 2062
      |||||||
Db 2162 TCTGTAATAACGTAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2221
Qy 2063 tgaacacctgtatgagaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaa 2118
      |||||||
Db 2222 TGAAACCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2277

RESULT 10
AX079372 274 bp DNA PAT 22-FEB-2001
LOCUS Sequence 116 from Patent WO0107611.
DEFINITION AX079372
ACCESSION AX079372
VERSION AX079372.1 GI:13158932
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 274)
AUTHORS Baker,K.P., Goddard,A. and Wood,W.I.
TITLE Human polypeptides and methods for the use thereof
JOURNAL Patent: WO 0107611-A 116 01-FEB-2001;
Genentech, Inc. (US)
PUBS location/Qualifiers
SOURCE 1. 274
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 79 a 53 c 52 g 84 t 6 others
ORIGIN

```

```

Query Match 12.28; Score 258.4; DB 6; Length 274;
Best Local Similarity 95.68; Pred. No. 3.3e-58;
Matches 262; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```

```

Qy 238 gtaattcccccgtctacatactctgacacaatctcatgaagaagattgagatgtat 297
      |||||||
Db 1 GTCATTTCCCGCGCTTTATATCTGTAACACATTTTCATGAAAGATTCCACATGTTAT 60
Qy 298 gggctgtagccaaaagctagaagaataagacaaatatlgtggaacacacataaagt 357
      |||||||
Db 61 GGCGTGAAGCCAAAAGGATGAGAGATTAAGACAAATATGTGGAAGCACACATATAAGT 120
Qy 358 ttctcaactcctaccaggagatgagacatttgaaacagctccggccaagaagtcacca 417
      |||||||
Db 121 TTATATCAANTTCATACCGGAGATGAGAGATTGAGACAGTTCCGCAAGANGTCACCA 180
Qy 418 agtctcttccagtgatattctcattccacttctgcaacttctgctctctc 477
      |||||||
Db 181 AGTGATTTTCTAGTATATTTTCATTCACCTTTTGCCCAANTACTTGTTTTCG 240

```

```

Qy 478 taatgtaacctgttccacagcaactactatcag 511
      |||||||
Db 241 TATATGTAACCTGTTTCCACAGCAATATANTATGATGAG 274

```

```

RESULT 11
AC023648 180268 bp DNA HTG 17-MAY-2000
LOCUS Homo sapiens chromosome 12 clone RP11-624H4 map 12, WORKING DRAFT
DEFINITION SEQUENCE, 35 unordered pieces.
AC023648
AC023648.3 GI:7884355
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 180268)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 12, clone RP11-624H4
JOURNAL Unpublished
PUBS 2 (bases 1 to 180268)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Bouhgalter,B., Brown,A., Burkett,G., Campoliano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Deatellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Glinde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heathford,A., Horton,L.,
Howard,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Laroque,K., Lechoczky,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McKernan,K., McPheters,R., Meldrum,J.,
Menus,L., Mihova,T., Miranda,C., Mieng,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,
Peterson,K., Pierre,N., Pisanic,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Testfay,S., Theodore,J., Tirrell,A.,
Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.

```

```

TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 16516
Center clone name: 624.H.4
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 153867 bases at least Q40
Consensus quality: 164187 bases at least Q30
Consensus quality: 170599 bases at least Q20
Insert size: 176000; agarose-fp
Quality coverage: 5.3 in Q20 bases; agarose-fp
Quality coverage: 5.3 in Q20 bases; sum-of-contrigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 35 contrigs. The true order of the pieces

```



Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,  
 Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korah, J.,  
 Kovar, C., Krawiec, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C.,  
 Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,  
 Lounsgaard, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,  
 Martinez, E., Massey, E., Mawhinney, E., McLeod, M. P., Meador, M.,  
 Mel, G., Metker, M., Miner, G., Miner, Z., Mitchell, J., Mohabati, K.,  
 Morgan, M., Morris, S., Moser, M., Neal, D., Mitchell, J., Newton, N.,  
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokoko, S.,  
 Ogunu, G., Okunnu, G., Orgunye, N., Oviedo, R., Pace, A., Payton, B.,  
 Peery, J., Perez, L., Peters, L., Plickens, R., Primus, E., Pull, L.,  
 Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M.,  
 Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoochitari, N.,  
 Stilson, I., Sodergren, E., Sonalle, T., Sparks, A., Stanley, H.,  
 Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A.,  
 Tang, H., Tang, J., Taylor, C., Taylor, T., Tellirod, B., Thomas, N.,  
 Thomas, S., Usmani, K., Vasquez, J., Ward, V., Villalona, D., Vinson, R.,  
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
 Wellington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S.,  
 Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstock, G., and Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 336508)  
 Worley, K.C.  
 Direct Submission  
 Submitted (08-OCT-1999) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jun 25, 2001 this sequence version replaced gi:14190620.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 -----  
 Center project name: HMLJ  
 Center clone name: RP11-392116  
 ----- Summary Statistics

```

Chemistry: Dye-terminator, model
Chemistry: Dye-terminator Big Dye: 86% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 318894 bases at least Q40
Consensus quality: 342195 bases at least Q30
Consensus quality: 354034 bases at least Q20
Estimated insert size: 350560; sum-of-ctrls estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 3.4x in Q20 bases; sum-of-ctrls estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_drift\_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It is current!

```

```

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_drift\_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 49 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 34038: contig of 34038 bp in length
* * 34039 34138: gap of unknown length
* * 34139 34138: gap of unknown length
* * 63289 63289: contig of 29151 bp in length
* * 63290 63390: gap of unknown length
* * 63390 81254: contig of 17865 bp in length
* * 81255 81354: gap of unknown length
* * 81355 97046: contig of 15692 bp in length
* * 97047 97146: gap of unknown length
* * 97147 109167: contig of 12021 bp in length
* * 109168 109267: gap of unknown length

```



DB 304434 GAAGTAAATCGG 304422

# RESULT 13

AC011600 LOCUS AC011600 336508 bp DNA HTG 27-JUN-2001  
 DEFINITION Homo sapiens chromosome 12 clone RP11-392116, WORKING DRAFT  
 SEQUENCE, 49 unordered pieces.  
 AC011600  
 AC011600 28 GI:14547729  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 AUTHORS 1 (bases 1 to 336508)

Muzny, D.N., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.L., Amaral, H.C., Are, J.R., Banks, T., Barbarella, J., Benton, J., Bimberg, K., Blankenburg, R., Bonnin, D., Bouck, J., Bowie, S., Brice, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy, Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homs, P., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Lousseng, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Meshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabab, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokweto, S., Oguh, M., Okunolu, G., Oreguene, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shooshbari, N., Stinson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Thomas, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wlarczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Unpublished 2 (bases 1 to 336508)  
 Direct Submission  
 Submitted (08-OCT-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Jun 25, 2001 this sequence version replaced gi:14190620.

## COMMENT

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: HMLU  
 Center clone name: RP11-392116

----- Summary Statistics  
 Sequencing vector: Plasmid; M77789  
 Sequencing vector: M13; 108821  
 Chemistry: Dye-terminator Big Dye 88% of reads  
 Chemistry: Dye-terminator Big Dye 88% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 318894 bases at least Q40  
 Consensus quality: 342195 bases at least Q30  
 Consensus quality: 354034 bases at least Q20  
 Estimated insert size: 350560, sum-of-contigs estimation  
 Quality coverage: 0x in Q20 bases; agarose-TP estimation  
 Quality coverage: 3.4x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_drift\\_date.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_drift_date.html))  
 \* NOTE: This sequence may represent more than one clone.  
 \* NOTE: This is a "working draft" sequence. It currently  
 \* consists of 49 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1  
 \* 34038: contig of 34038 bp in length  
 \* 34039  
 \* 34138: gap of unknown length  
 \* 34139  
 \* 63289: contig of 29151 bp in length  
 \* 63290  
 \* 63390: gap of unknown length  
 \* 81254: contig of 17865 bp in length  
 \* 81255  
 \* 81354: gap of unknown length  
 \* 81355  
 \* 97046: contig of 15992 bp in length  
 \* 97047  
 \* 97146: gap of unknown length  
 \* 97147  
 \* 109167: contig of 10201 bp in length  
 \* 109168  
 \* 109267: gap of unknown length  
 \* 109268  
 \* 123513: contig of 14646 bp in length  
 \* 123914  
 \* 124013: gap of unknown length  
 \* 124014  
 \* 133886: contig of 9873 bp in length  
 \* 133887  
 \* 133986: gap of unknown length  
 \* 133987  
 \* 144693: contig of 10707 bp in length  
 \* 144694  
 \* 144793: gap of unknown length  
 \* 144794  
 \* 154244: contig of 9451 bp in length  
 \* 154245  
 \* 154344: gap of unknown length  
 \* 154345  
 \* 163322: contig of 8978 bp in length  
 \* 163323  
 \* 163422: gap of unknown length  
 \* 163423  
 \* 172826: contig of 9404 bp in length  
 \* 172827  
 \* 172926: gap of unknown length  
 \* 172927  
 \* 180526: contig of 7600 bp in length  
 \* 180527  
 \* 180626: gap of unknown length  
 \* 180627  
 \* 188940: contig of 8314 bp in length  
 \* 188941  
 \* 189040: gap of unknown length  
 \* 189041  
 \* 195336: contig of 6596 bp in length  
 \* 195337  
 \* 195736: gap of unknown length  
 \* 195737  
 \* 201841: contig of 6105 bp in length  
 \* 201842  
 \* 201941: gap of unknown length  
 \* 201942  
 \* 209588: contig of 7647 bp in length  
 \* 209589  
 \* 209688: gap of unknown length  
 \* 209689  
 \* 217437: contig of 7749 bp in length  
 \* 217438  
 \* 217537: gap of unknown length  
 \* 217538  
 \* 222495: contig of 4558 bp in length  
 \* 222496  
 \* 222595: gap of unknown length  
 \* 222596  
 \* 227782: contig of 5187 bp in length  
 \* 227783  
 \* 227882: gap of unknown length  
 \* 227883  
 \* 233115: contig of 5233 bp in length  
 \* 233116  
 \* 233215: gap of unknown length  
 \* 233216  
 \* 238189: contig of 4974 bp in length  
 \* 238190  
 \* 238289: gap of unknown length  
 \* 238290  
 \* 244865: contig of 6576 bp in length  
 \* 244866  
 \* 244965: gap of unknown length  
 \* 244966  
 \* 249984: contig of 5019 bp in length  
 \* 249985  
 \* 250084: gap of unknown length  
 \* 250085  
 \* 254445: contig of 4361 bp in length  
 \* 254446  
 \* 254545: gap of unknown length  
 \* 259133: contig of 4588 bp in length

*	25913.4	25923.3	gap of unknown length
*	25922.4	264486:	contig of 5253 bp in length
*	264487	264586:	gap of unknown length
*	264587	267850:	contig of 3264 bp in length
*	267851	267950:	gap of unknown length
*	267951	271465:	contig of 3515 bp in length
*	271466	271565:	gap of unknown length
*	271566	271968:	contig of 3403 bp in length
*	271969	275068:	gap of unknown length
*	275069	278811:	contig of 3743 bp in length
*	278812	278911:	gap of unknown length
*	278912	281845:	contig of 2934 bp in length
*	281846	281945:	gap of unknown length
*	281946	285805:	contig of 3860 bp in length
*	285806	285905:	gap of unknown length
*	285906	288213:	contig of 3308 bp in length
*	288214	288313:	gap of unknown length
*	288314	293460:	contig of 4147 bp in length
*	293461	293560:	gap of unknown length
*	293561	297333:	contig of 3773 bp in length
*	297334	297433:	gap of unknown length
*	297434	300053:	contig of 2626 bp in length
*	300060	300159:	gap of unknown length
*	300160	302533:	contig of 2374 bp in length
*	302534	302633:	gap of unknown length
*	302634	306196:	contig of 3563 bp in length
*	306197	306296:	gap of unknown length
*	306297	309165:	contig of 2869 bp in length
*	309166	309266:	gap of unknown length
*	309266	311791:	contig of 2526 bp in length
*	311792	311891:	gap of unknown length
*	311892	315049:	contig of 3158 bp in length
*	315050	315150:	gap of unknown length
*	315150	317892:	contig of 2643 bp in length
*	317893	317892:	gap of unknown length
*	317893	320763:	contig of 2871 bp in length
*	320764	322189:	contig of 2326 bp in length
*	322189	322289:	gap of unknown length
*	322290	325971:	contig of 2662 bp in length
*	325972	326071:	gap of unknown length
*	326072	328644:	contig of 2570 bp in length
*	328644	328741:	gap of unknown length
*	328742	331416:	contig of 2675 bp in length
*	331417	331516:	gap of unknown length
*	331517	334019:	contig of 2503 bp in length
*	334020	334119:	gap of unknown length
*	334120	336508:	contig of 2369 bp in length

FEATURES	Location/Qualifiers
source	1..36508
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="12"
	/clone="RP11-392116"
BASE COUNT	104868 a 63616 c 63734 g 99395 t
ORIGIN	4895 others

Query Match	7.2%;	Score 152.8;	DB 2;	Length 336508;
Best Local Similarity	98.7%;	Pred. No. 1.8e-29;		
Matches 154; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0

Qy 128 tcggtcttcaaaagcttcacctttccccaagaagcagatglgaagaactgatgcatagtc 187  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 235359 TAGGTCTTCAAAAGCTTCACCTTTCTCCAAAAGGAGATGTGAAGAAGACTGTATNCTCATATGT 235418

Qy 188 ggtcaaccagacaaaagcgattaatgggaataacacatcglttccttgygfcgtcaattccc 247  
|||||  
Db 235419 GGTACCAAGACAAAAGCGATTAAATGGGAATTCACATCGTTCCTTGGGTCGTCAATTTCCC 235478

Qy 248 ccgcctctatactctgttacacaatcctcatgaag 283  
|||||  
Db 235479 ccgcctctatgtcctgtacacacatcttcatagaag 235514

RESULT	14
LOCUS	AC023070/c
DEFINITION	Mus musculus clone RP23-11118, WORKING DRAFT SEQUENCE, 14 unordered pieces.
ACCESSION	AC023070
VERSION	AC023070.5 GI:13940736
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FUZZTOP.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
AUTHORS	Barren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE	Mus musculus, clone Rp23-11118
JOURNAL	unpublished
REFERENCE	2 (bases 1 to 212282)
AUTHORS	Barren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

TITLE Direct Submission  
JOURNAL Submitted (08-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 370 Charles Street, Cambridge, MA 02141, USA  
COMMENT On May 4, 2001 this sequence version replaced g1:13123357.

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L5130
Center clone name: 11-T18
-----
Summary Statistics
Sequencing vector: M13; M77815; 37% of reads
Sequencing vector: Plasmid; n/a; 63% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 208523 bases at least Q40
Consensus quality: 209672 bases at least Q20
Consensus quality: 210447 bases at least Q20
Insert size: 198000; agarose-fp
Insert size: 210982; sum-of-contigs
Quality coverage: 8.4 in Q20 bases; agarose-fp
Quality coverage: 7.9 in Q20 ba.
NOTE: This is a "working draft" sequence. It currently
consists of 14 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1
2135: contig of 2135 bp in length
2136 2235: gap of 100 bp

```

*	2236	3382:	contig of 1147 bp in length
*	3383	3482:	gap of 100 bp
*	3483	61437:	contig of 57955 bp in length
*	61438	61537:	gap of 100 bp
*	61538	66416:	contig of 4879 bp in length
*	66417	66516:	gap of 100 bp
*	66517	72390:	contig of 5874 bp in length
*	72391	72490:	gap of 100 bp
*	72491	79436:	contig of 6946 bp in length
*	79437	79536:	gap of 100 bp
*	79537	86074:	contig of 6538 bp in length
*	86075	86174:	gap of 100 bp
*	86175	96577:	contig of 10403 bp in length
*	96578	96677:	gap of 100 bp
*	96678	111853:	contig of 15176 bp in length
*	111854	111953:	gap of 100 bp
*	111954	129092:	contig of 17133 bp in length
*	129093	129192:	gap of 100 bp
*	129193	152023:	contig of 22831 bp in length
*	152024	152123:	gap of 100 bp
*	152124	179718:	contig of 27595 bp in length
*	179719	179818:	gap of 100 bp
*	179819	208560:	contig of 28742 bp in length
*	208561	208660:	gap of 100 bp
*	208661	212282:	contig of 3622 bp in length.

Location/Qualifiers

```
misc_feature 1. .2135
```

```
misc_feature    2236.  .3382
```

```
misc_feature    3483.    .61437
```

```
misc_feature    61538.66416
```

```
misc_feature 66517.  .72390
```

misc_feature	72491.	.79436
72491.	.79436	

misc feature 795337, 86074

```
misc feature 86175.96577 /note-assembly-11
```

```
misc feature 96678. 111853 /noce-assembly-11
```

```
misc feature
111954 129092 .
/mole=assembly_fragment
```

```

/note="assembly_fragment
130103 153033

```

```

/note="assembly_fragment
153134 170718

```

```

/note="assembly_fragment
170010 200560

```

```
/note="assembly_fragment"
```

	a	c	g	t	others
BASE COUNT	53829	50391	49748	56994	1320
ORIGIN			.		

Query Match	7.28;	Score 152.6;	DB 2;	Length 212282;
Best Local Similarity	63.78;	Pred. No. 1.9e-29;		
Matches 319;	Conservative 0;	Mismatches 164;	Indels 18;	Gaps 5

Qy 1583 ccaccccttctatataagcacaacaaacacagltgttcttcttctgtcaccctcactgcga 1642

Db	50601	CCATCCCTTTCGGTGGAGCAACAAATAGGCGTGTTCCTCCGCACTATCATGGCGCTAA	50542
Qy	1543	tgggtcaaccactcctcttlltccaaccctctgaacaccttllctctgaaggtaactgcag	1702
Db	50541	CTTCCACCCCTCCCTCGATATTCTATGGACACTTTTTTTTTTAAATTAATTAATGG	50482
Qy	1703	gaatattctatcttgaaagaatatacagagagtaacagtgactctctgaacccttcctcaatg	1762
Db	50481	GAACTATTCTAAAGAGAGGCGGT-----GAGAAAGAATGGCTGGCGCCCTCCCTCGT	50427
Qy	1763	ttctcagccttaactctccatgtctcttllctcgggctgtatacagccctctgtgatact	1822
Db	50426	TGTC-----CCAGTGTTCATTCTTCAGGCGCTGCAACAGCCCTTAAGGACCTTA	50376
Qy	1823	caactctgctgcctccactctgtatcagcagctcaactgaactgaactgagctgcttc	1882
Db	50375	CAAGTTTGGACTCCACAGCTTGTCACTAAGCCATCTGTAACAGAGAGGCGTGGCTTC	50316
Qy	1883	tctggccatgatgc--acactctaaagtaactaactaactgcccagcctcggggagatcag	1940
Db	50315	TCTAAGCGCTGGGTCCAAAACCAACAGATCTAATTACTGCCAGCCCTGGGAGACAG	50256
Qy	1941	gagaagctcgcacagtaagtaagcttgggctttagccttctgtgtgcatacagtaactaga	2000
Db	50255	GAGC--TTTGCATAGATGATGAGTTGGGCTTAACCTGCTGCTGTGACATCTATCCCTGTC	50197
Qy	2001	gtctctgaataactatctgtlaaagtcaltgaagcaactgttlltaaacccaagtaagaagt	2060
Db	50196	GTTTGTGAATAAATCTACTTAATGCAAGAAGCACTG--GTTTAAGCCTAAGTAAAGACTG	50138
Qy	2061	cttgaaacctgtgatgagaa	2081
Db	50137	CTGTAAACCGATTTCGGGAA	50117

Db 50137 CCTGAACCCAGTTTCTGGGA 50117

## RESULT 15

LOCUS

ACCESSION

## KEYWORDS

1  
2  
3  
4  
5  
6

## REFERENCE

EVOLUTION

TITLE

JOURNAL  
OF  
BREEDING

## AUTHORS

TITLE

COMMENT

## FEATURES

/organism="Homo sapiens"

```

/db xref="taxon:9606"
/cell_line="KATO III"
/cell_type="signet-ring cell carcinoma"
/clone="KAT11312"
/clone_id="KAT"
/note="cloning vector PME18FL3"
261. 962

```

```
/cell_line="KATO III"  
/cell_type="signal-ri
```

```
/cell_type="signet-ring cell carcinoma"
/closure="KAT11213"
```

```
/clone="KAT11312
```

```
/clone_lib="KAT"
```

```

7 more cloning vector pME185FL3
261. .962

```

201. 962  
/note="un

```
/note="unnamed protein product"  
/codon_start=1
```

```
/codon_start=1
/protein_id="PAB
```

```
/protein_id="BAB15514.1"
/db_xref="GT:10439536"
```

```
/db_xref="GI:10439526"
```

EAKKRWETKSNMGYM"  
368 C 450 a 603 +

a	368 c	450 g	603 t
---	-------	-------	-------

BASE COUNT	736 a	368 c	450 g	603 t
ORIGIN				

ORIGIN

Query Match

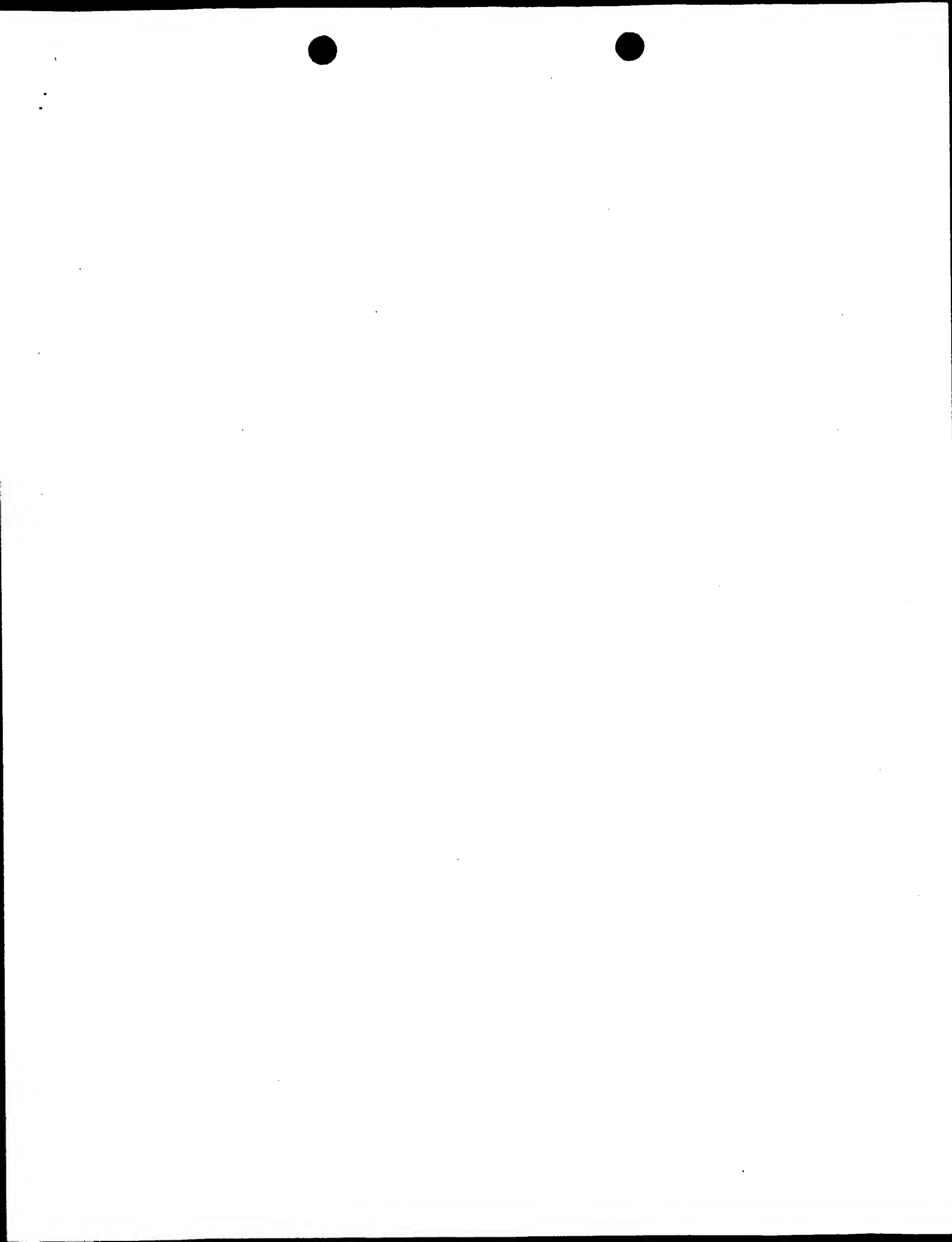
Query match	2.98;	Score 60.4;	DB 9;	Length 2157;
Best Local Similarity	71.88;	Pred. No. 4.4e-05;		

Best Local Similarity	71.8%;	Pred. No. 4.4e-05;
Matches	79;	Conservative 0;
		Mismatches 31

```
Matches      /9;  Conservative      0;  Mismatches      31;  Indels      0;  Gaps      0;
```

[illegible][illegible]

Search completed: December 20, 2001, 19:18:59  
Job time: 9751 sec







Db	1561	gcagagcacttgagggtgtttccctccacccttcttcgataatgacccaaaaatcaagatgtgtttg	1620
Qy	1621	tttttttgcaaccttcacatgvcagatgggtctaaccaactctttltaaacccctctgaacact	1680
Db	1621	tttttttgcaaccttcacatgvcagatgggtctaaccaactctttltaaacccctctgaacact	1680
Qy	1681	ttttctgatgggtaaccttgcagaagataattcttatgttgaaaagtaacagaagatcaagtg	1740
Db	1681	ttttctgatgggtaaccttgcagaagataattcttatgttgaaaagtaacagaagatcaagtg	1740
Qy	1741	ctctcttaaccccttcctccaaatgtttcttaagccttcaactcccaattctcttctctggcgt	1800
Db	1741	ctctcttaaccccttcctccaaatgtttcttaagccttcaactcccaattctcttctggcgt	1800
Qy	1801	attacagccctctgtgataltcttaactctgtcgtccctccactgtatgtgacagatccaact	1860
Db	1801	attacagccctctgtgataltcttaactctgtcgtccctccactgtatgtgacagatccaact	1860
Qy	1861	gtaacctgaacagatggctcgtccctctctctctgtggccaatgcatcaacactgtgaagttactaat	1920
Db	1861	gtaacctgaacagatggctcgtccctctctctctgtggccaatgcatcaacactgtgaagttactaat	1920
Qy	1921	gccacagccttgaggagatccagagaaggtctgcgaatgtaagttaggttttagctttgt	1980
Db	1921	gccacagccttgaggagatccagagaaggtctgcgaatgtaagttaggttttagctttgt	1980
Qy	1981	gtgtgcatacagtgactataaggtctctgtaataactcttbtlaaatgcatgtgaagcatgttt	2040
Db	1981	gtgtgcatacagtgactataaggtctctgtaataactcttbtlaaatgcatgtgaagcatgttt	2040
Qy	2041	tttaaacccaagttaagactgtcttgaaacctgttgalatgtaaaaaaanaaaaaaanaaaa	2100
Db	2041	tttaaacccaagttaagactgtcttgaaacctgttgalatgtaaaaaaanaaaaaaanaaaa	2100
Qy	2101	aaaaaaaaaaaaaaaaaaaaa 2118	
Db	2101	aaaaaaaaaaaaaaaaaaaaa 2118	

XX	AAH45403	standard; cDNA; 2002 BP.
XX	AAH45403;	
DT	05-SEP-2001	(first entry)
XX		
DE	Human cDNA encoding ribosomal protein L14.22.	
XX		
KW	Human; ribosomal protein L14.22; malignant neoplasm; haemopathy;	
KW	HIV infection; immunological disease; inflammatory disorder; cytostatic;	
KW	haemostatic; virucide; immunomodulatory; antiinflammatory; ss.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	413..1024
FT		/*tag= a
FT		/product= "Ribosomal protein L14.22"
XX		
PN	WO200138389-A1.	
XX		
PD	31-MAY-2001.	
XX		
PF	20-NOV-2000; 2000WO-CN00471.	
XX		
PR	24-NOV-1999; 99CN-0124098.	
XX		
PA	(BIOR-) BIOROAD GENE DEV LTD SHANGHAI.	
XX		
PI	Mao Y, Xie Y;	
XX		
DR	WPI; 2001-355906/37.	

DR	P-PSDB; AAG62443.
xx	
PT	Human ribosomal protein L14.22 and encoded polynucleotide, used in
PT	diagnosis and treatment of malignant tumors, hemopathy, human
PT	immunodeficiency virus infection, immunological diseases and
PT	inflammation -
xx	
PS	Claim 6; Page 19-20; 30pp; Chinese.
xx	
CC	This invention relates to human ribosomal protein L14.22 and cDNA
CC	encoding it. The invention includes a vector containing the cDNA, a host
CC	cell transformed with the vector and an antibody directed against human
CC	ribosomal protein L14.22. The cDNA and protein sequences are used in the
CC	diagnosis and treatment of malignant neoplasms, haemopathy, human
CC	immunodeficiency virus (HIV) infection, immunological diseases and
CC	various inflammatory disorders. Use of the protein and nucleotide
CC	sequence may result in cytosstatic; haemostatic; virocidic;
CC	immunomodulatory and antiinflammatory activity. This sequence represents
CC	cDNA encoding human ribosomal protein L14.22.
xx	
SO	Sequence 2002 BP; 456 A; 494 C; 467 G; 585 T; 0 other;

Query Match	85.88;	Score 1817.4;	DB 22;	Length 2002;
Best Local Similarity	94.18;	Pred. No. 0;		
Matches 1947;	Conservative	0;	Mismatches 6;	Indels 116;
				Gaps 1;

QY	1	ctgtgaagaatgagcctctccagaggtgtgctgtggtctgtgcgtgtgtgtgtggtctggcag	60
Db	50	ctgtgaagaatgagcctctccagaggtgtgctgtggtctgtgcgtgtgtgtgtggtctggcag	109
QY	61	tcaaccctggagcatcttctaccccgagagctgcacaattgtgcgtctcgtgcctgtctgag	120
Db	110	tcaaccctggagcatcttctaccccgagagctgcacaattgtgcgtctcgtgcctgtctgag	169
QY	121	gggcccctcgtgcttcaaaagctcaaccttctccaagggcagatgtgaagaacttgat	180
Db	170	gggcccctcgtgcttcaaaagctcaaccttctccaagggcagatgtgaagaacttgat	229
QY	181	cttatgtgtgtaaccaagcaaaagcgtatlaatgtggaataaccatcgctttcttggtctgc	240
Db	230	cttatgtgtgtaaccaagcaaaagcgtatlaatgtggaataaccatcgctttcttggtctgc	289
QY	241	attcccccgcctctcatctctctgtacacaactcttcacgaaaggaatgcagatgttatcgg	300
Db	290	attcccccgcctctcatctctctgtacacaactcttcacgaaaggaatgcagatgttatcgg	329
QY	301	ctgatgccaanaaaggctagaagaataaagacaataatgtgtgaagcacacataaagtctc	360
Db	330	-----	329
QY	361	atcaactlccataccggagatgtgagcatcttgaagacgattccgcgaagaagccaag	420
Db	330	-----	353
QY	421	gctcttccatagtaatttccatctcacctcttgcacaactactgctctcttgtaa	480
Db	354	gctcttccatagtaatttccatctcacctcttgcacaactactgctctcttgtaa	413
QY	481	tgtacctgtttccagagcaactactgatacaggcatcttctgagccccaacaacaacatg	540
Db	414	tgtacctgtttccagagcaactactgatacaggcatcttctgagccccaacaacaacatg	473
QY	541	attctcttgatactatcatatgctttcttcggagagctccaccaccgaataattatgatt	600
Db	474	attctcttgatactatcatatgctttcttcggagagctccaccaccgaataattatgatt	533
QY	601	tagaagaagtcataccctctcatcttctgtatgagaagactcggctgtgcgtctgaacagatctg	660
Db	534	tagaagaagtcataccctctcatcttctgtatgagaagactcggctgtgcgtctgaacagatctg	593
QY	661	gcaccaagaatacagcgtgtgtaaccaaccagacaatacatgatatacttggctctgaagagt	720

Db 594 gcaaccaagatacagctggtgtaaccaccagcaatataatgatacttggtctgagagagt 653  
 QY 721 gttctcacaacatcctctgggagcaatgaacaccacagagcttgcaagtgaagccttga 780  
 Db 654 gttctcacaacatcctctgggagcaatgaacaccacagagcttgcaagtgaagccttga 713  
 QY 781 gccgggacatgcttcacatcttaactgctcctcctcttgtagaatacgtttgaaga 840  
 Db 714 gccgggacatgcttcacatcttaactgctcctcctcttgtagaatacgtttgaaga 773  
 QY 841 ctctaacacatgcttattacacacatgagcaaggtttggcaaaactgaggatctggcagc 900  
 Db 774 ctctaacacatgcttattacacacatgagcaaggtttggcaaaactgaggatctggcagc 833  
 QY 901 tgaactgctcagaagataaatacaggttgatctcccgctggcctgaattctcaacagatg 960  
 Db 834 tgaactgctcagaagataaatacaggttgatctcccgctggcctgaattctcaacagatg 893  
 QY 961 gtgaagaataggtgtcgaactgtgctggagaaatggtcgaagattctcgcagcctgaag 1020  
 Db 894 gtgaagaataggtgtcgaactgtgctggagaaatggtcgaagattctcgcagcctgaag 953  
 QY 1021 aagctgagctgtctctctctgtgtgacaaagtgtgctcctcctcaacacatccttggga 1080  
 Db 954 aagctgagctgtctctctctgtgtgacaaagtgtgctcctcctcaacacatccttggga 1013  
 QY 1081 caaggcgtcgaatgaacacatgagcagatgcatgtctcctgcaatctgtatagataagcag 1140  
 Db 1014 caaggcgtcgaatgaacacatgagcagatgcatgtctcctgcaatctgtatagataagcag 1073  
 QY 1141 tgcaggaacaaacagacacttgccagcaaaagtctgtgtactgttaagtgtgtgtggagc 1200  
 Db 1074 tgcaggaacaaacagacacttgccagcaaaagtctgtgtactgttaagtgtgtgtggagc 1133  
 QY 1201 agagaagagagcagagggcagcttgcacagccttgccagcaacactgtgtggagcagaca 1260  
 Db 1134 agagaagagagcagagggcagcttgcacagccttgccagcaacactgtgtggagcagaca 1193  
 QY 1261 ttcctcctcaacagcctcagacatgaacaaacccctctgtctagaggtgtgtgtgtgtgt 1320  
 Db 1194 ttcctcctcaacagcctcagacatgaacaaacccctctgtctagaggtgtgtgtgtgtgt 1253  
 QY 1321 catctgtctccctcctcaataataataatagctggaactgagcagcctcactgtggcctt 1380  
 Db 1254 catctgtctccctcctcaataataataatagctggaactgagcagcctcactgtggcctt 1313  
 QY 1381 taactgtatgtgtcagttatcagttatcagttatcagttatcagttatcagttatcagtt 1440  
 Db 1314 taactgtatgtgtcagttatcagttatcagttatcagttatcagttatcagttatcagtt 1373  
 QY 1441 taattgctctagagcctatcagcctcactcacttacttaactacatctgtgtgtgtgtgt 1500  
 Db 1374 taattgctctagagcctatcagcctcactcacttacttaactacatctgtgtgtgtgtgt 1433  
 QY 1501 tgttagaataattttgt 1560  
 Db 1434 tgttagaataattttgt 1493  
 QY 1561 gcatgtaactgt 1620  
 Db 1494 gcatgtaactgt 1553  
 QY 1621 tttttgtcactcactcactcactcactcactcactcactcactcactcactcactcactcact 1680  
 Db 1554 tttttgtcactcactcactcactcactcactcactcactcactcactcactcactcactcact 1613  
 QY 1681 tttctgtatgt 1740  
 Db 1614 tttctgtatgt 1673  
 QY 1741 ctcttgacccctcctcctcaatgtttctagcctcactcactcactcactcactcactcactcact 1800  
 Db 1674 ctcttgacccctcctcctcaatgtttctagcctcactcactcactcactcactcactcactcact 1733

QY 1801 attacagccctctgtgtatcttcaactctgtgctgctcactgtgtagcagcagctcaact 1860  
 Db 1734 attacagccctctgtgtatcttcaactctgtgctgctcactgtgtagcagcagctcaact 1793  
 QY 1861 gtaactgacagtggtcgtcctctctctgtggcactgataacacactgtgaagtaacttaact 1920  
 Db 1794 gtaactgacagtggtcgtcctctctctgtggcactgataacacactgtgaagtaacttaact 1853  
 QY 1921 gcccaagcctgggagatcagagagaggtctgtcagatgtagtaagttgggttgccttgt 1980  
 Db 1854 gcccaagcctgggagatcagagagaggtctgtcagatgtagtaagttgggttgccttgt 1913  
 QY 1981 gtgtgcatcagtgacttaagatctgttaataacttaattgaatcagtaagcactgttt 2040  
 Db 1914 gtgtgcatcagtgacttaagatctgttaataacttaattgaatcagtaagcactgttt 1973  
 QY 2041 ttaaacccaagtaagactgtgtgaacc 2069  
 Db 1974 ttaaacccaagtaagactgtgtgaagc 2002

RESULT 3  
 AAF18148  
 ID AAF18148 standard; DNA; 1645 BP.  
 AC AAF18148;  
 XX  
 DT 14-MAR-2001 (first entry)  
 XX  
 DE Lung cancer associated polynucleotide sequence SEQ ID 167.  
 DE  
 KW Human; lung cancer associated protein; neuroprotective; cytostatic;  
 KW cardioactive; immunomodulatory; muscular active; vulnerary;  
 KW gastrointestinal; nephrotoxic; antilinfecive; gynecological;  
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;  
 KW proliferative disorder; wound healing; infectious disease; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO20005180-A2.  
 XX  
 XX 21-SEP-2000.  
 XX  
 PD 08-MAR-2000; 2000MO-US05918.  
 PF  
 PR 12-MAR-1999; 99US-0124270.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (ROSE/) ROSEN C A.  
 PA XX  
 PI Ruben SM;  
 XX  
 DR WPI: 2000-587514/55.  
 DR P-PSDB: AAB58272.  
 PT Lung cancer associated gene sequences, referred to as lung cancer  
 PT antigens, useful for treatment, prevention, and diagnosis of disorders  
 PT such as lung cancer -  
 PS  
 PS Claim 1; Page 631; 1425pp; English.  
 CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer  
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer  
 CC associated proteins and polynucleotide sequences, their agonists, and  
 CC antagonists may have neuroprotective; cytoskeletal; cardioactive;  
 CC immunomodulatory; muscular active general; vulnerary; gastrointestinal  
 CC general; nephrotoxic; antilinfecive; gynecological; or antibacterial  
 CC activity. The invention also includes antibodies specific for the  
 CC protein or polynucleotide sequences. The lung cancer associated  
 CC polynucleotide sequences may be used for detection of lung cancer,  
 CC chromosome identification, as chromosome markers, and for numerous other  
 CC diagnostic or research purposes. The proteins may be used to treat

Db	1136	aagtgttttgtttttttgttaaccttaactggaatgcagggctaaaccaattcttttcaaccc	1195		
Qy	1670	tctgaacaactttcttcgtatgvtggttaacttcaggaatatcttatgtgaaaaagataacag	1729		
Db	1196	tctgaacaacttttctctgatgtaacttcaggaatatcttatgtgaaaaagataacag	1255		
Qy	1730	aagtaacaagtctcttcgttaacccttcctcaaatgtttctagccttcaactccactgtctt	1789		
Db	1256	aagtaacaagtctcttcgttaacccttcctcaaatgtttctagccttcaactccactgtctt	1315		
Qy	1790	tctctgagctctgatatacaagccctctgttgatcttcaactctgtctgtccctcactgtatga	1849		
Db	1316	tctctgagctctgatatacaagccctctgttgatcttcaactctgtctgtccctcactgtatga	1375		
Qy	1850	gcagtcacaactgttaactgaacagtgtgcctctctctctgtggccaatgatacacacctgtaag	1909		
Db	1376	gcagtcacaactgttaactgaacagtgtgcctctctctctgtggccaatgatacacacctgtaag	1435		
Qy	1910	tacttaattacgtcccaagcctgtggagatcagaagaagctctgcatagtatgaatggttgggt	1969		
Db	1436	tacttaattacgtcccaagcctgtggagatcagaagaagctctgcatagtatgaatggttgggt	1495		
Qy	1970	ttagcttttgtgtgtgatacagtaacttaagatggtctctgaataacttatgttlaaalgatg	2029		
Db	1496	ttagcttttgtgtgtgatacagtaacttaagatggtctctgaataacttatgttlaaalgatg	1555		
Qy	2030	aagcactgtttttaaaccacaagtaaaagactgtctgaaacctgttgaatgtaaaaaaaa	2089		
Db	1556	aagcactgtttttaaaccacaagtaaaagactgtctgaaacctgttgaatgtaaaaaaaa	1615		
Qy	2090	aaaaaaaaaaaaa 2102			
Db	1616	aaaaaaaaaaaaa 1628			
RESULT 4					
ID	AAK30152	standard; DNA: 1197 BP.			
XX	AAK30152;				
AC	18-JUN-1999	(first entry)			
DT					
XX	Human secreted protein gene 8.				
DE					
XX	Human; secreted protein; cancer; tumour; developmental abnormality;				
XX	foeal deficiency; blood disorder; immune system disorder; inflammation;				
KW	autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;				
KW	schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;				
KW	atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;				
KW	digestive disorder; endocrine disorder; infection; AIDS; ss.				
OS	Homo sapiens.				
XX					
XX	WO9910363-A1.				
PN					
XX	04-MAR-1999.				
PD					
XX					
PF	27-AUG-1998;	98WO-US17709.			
XX					
XX	29-AUG-1997;	97US-0056271.			
PR	29-AUG-1997;	97US-0056073.			
PR	29-AUG-1997;	97US-0056247.			
PR	29-AUG-1997;	97US-0056270.			
XX					
PA	(HUMA-) HUMAN GENOME SCI INC.				
XX					
PI	Fan P, Kyaw H, Rosen CA, Ruben SM, Wei YF;				
XX					
XX	WPI: 1999-190585/16.				
DR	P-PSDB: AAY04300.				
XX					
PT	New isolated human genes and the secreted polypeptides they encode				

PT - useful for diagnosis and treatment of e.g. cancers, neurological disorders, immune diseases, inflammation or blood disorders

PS Claim 1; Page 136; 170pp; English.

SQ Sequence 1197 BP; 289 A; 270 C; 287 G; 345 T; 6 other;

Query Match	51.7%;	Score 1095.6;	DB 20;	Length 1197;
Post local similarity	00.7%	Score 1095.6;	DB 20;	Length 1197;

Matches 1101; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

Db	666	tcatttcgcgatacgaccacaaatcaagttgtttgttttttgcacccctcactgacg	725
Qy	1646	gctaaccaactctttttccaaacccctcgaaacccctttttctgattggttaacttcagga	1705
Db	726	gctaaccaactctttttccaaacccctcgaaacccctttttctgattggttaacttcagga	785
Qy	1706	tattctattggaagaagataacaggaagtacaaagtgcctcttgacccttcctcaatgttt	1765
Db	786	tattctattggaagaagataacaggaagtacaaagtgcctcttgacccttcctcaatgttt	845
Qy	1766	ctagcctcactccatctgctctttcttcggcctgtaatacagccctctgtgattctcaa	1825
Db	846	ctagcctcactccatctgctctttcttcggcctgtaatacagccctctgtgattctcaa	905
Qy	1826	ctctgcctgcctcccaactgtgatgcagcagctcccaactgaactgaacgttgctctctct	1885
Db	906	ctctgcctgcctcccaactgtgatgcagcagctcccaactgaactgaacgttgctctctct	965
Qy	1886	ggcgcatgatcacacactgtgaagtgactaaatactgccagccctggggagatcagagaag	1945
Db	966	ggcgcatgatcacacactgtgaagtgactaaatactgccagccctggggagatcagagaag	1025
Qy	1946	gtctgcagatgtagtaagctggggttaagctcttctgtgtgtgcatacagtgactagaattct	2005
Db	1026	gtctgcagatgtagtaagctggggttaagctcttctgtgtgtgcatacagtgactagaattct	1085
Qy	2006	gtaataactattgttaaatgcataggaagcactgtttttaaccagaataaagctgtctga	2065
Db	1086	gtaataactattgttaaatgcataggaagcactgtttttaaccagaataaagctgtctga	1145
Qy	2066	aacctgtgtgatcgtgaaaaaataaaaaaataaaaaa 2100	
Db	1146	aacctgtgtgatcgtgaaaaaataaaaaaataaaaaa 1180	

RESULT	5
AAH03869	
ID	AAH03869 standard; cDNA; 783 BP.
XX	
AC	AAH03869;
XX	
DT	26-JUN-2001 (first entry)
XX	
DE	Human cDNA clone (5'-primer) SEQ ID NO:704.
XX	
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX	
XX	Homo sapiens.
OS	
PN	EPI074617-A2.
XX	
PD	07-FEB-2001.
XX	
PF	28-JUL-2000; 2000EP-0116126.
XX	
PR	29-JUL-1999; 99JP-0248036.
XX	
PR	27-AUG-1999; 99JP-0300253.
PR	11-JAN-2000; 2000JP-0118776.
PR	02-MAY-2000; 2000JP-0183767.
PR	09-JUN-2000; 2000JP-0241899.
XX	
PA	(HELI-) HELIX RES INST.
XX	
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI	Isihii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX	
DR	WPI: 2001-318749/34.
XX	
PT	Primer sets for synthesizing polynucleotides, particularly the 5602
PT	full-length cDNAs defined in the specification, and for the detection
PT	and/or diagnosis of the abnormality of the proteins encoded by the
PT	full-length cDNAs .
XX	







OY 181 ctatgtgttaacccaagacaaagcattatggaatcaccatcgttcttgggtcgc 240  
 |||||  
 DB 204 ctatgtgttaacccaagacaaagcattatggaatcaccatcgttcttgggtcgc 263  
 OY 241 attccccgcctctatcctgtgtacacatcttcataagaa 284  
 |||||  
 DB 264 attccccgcctctatcctgtgtacacatcttcataagaa 307

## RESULT 10

AAF93295  
 ID AAF93295 standard; cDNA; 274 BP.

AC AAF93295;

DT 21-MAY-2001 (first entry)

DE Lung carcinoma cDNA encoding SRT protein SEQ ID 116.

XX Human; SRT; gene therapy; gene mapping; tissue typing; ss.

XX Homo sapiens.

PN WO200107611-A2.

PD 01-FEB-2001.

PF 21-JUL-2000; 2000MO-US20006.

PR 26-JUL-1999; 99US-0145701.

PA (GETH ) GENENTECH INC.

PI Baker KP, Goddard A, Wood WI;

XX WPI; 2001-112729/12.

PT New isolated nucleic acid molecule encoding a SRT polypeptide is useful  
 for production of recombinant SRT polypeptides, gene mapping,  
 PT diagnosing genetic disorders and for gene therapy -

PS Claim 2; Fig 116; 663pp; English.

XX Sequences AAF93180 - AAF93743 represent polynucleotide sequences encoding  
 CC human SRT proteins. The cDNA sequences are isolated from various  
 CC different human tissue cDNA libraries. The invention relates to a method  
 CC for detecting cDNA encoding an SRT protein, a vector containing cDNA  
 CC encoding SRT, a host cell transformed with the vector, an isolated SRT  
 CC polypeptide, and an antibody which binds to SRT. The polynucleotide  
 CC sequence can be used in gene therapy and is useful in the recombinant  
 CC production of SRT polypeptides, as a hybridisation probe to screen  
 CC libraries to isolate cDNAs with sequence identity to SRT polypeptides, to  
 CC map the gene encoding the SRT polypeptides and analysing genetic  
 CC disorders, tissue typing and disease tissue detection. The SRT  
 CC polynucleotide sequences can be used in polymerase chain reaction,  
 CC screening for new therapeutic molecules and generation of antisense RNA  
 CC and DNA.

SO Sequence 274 BP; 79 A; 53 C; 52 G; 84 T; 6 other:

Query Match 12.24; Score 258.4; DB 22; Length 274;

Best Local Similarity 95.6%; Pred. No. 2,4e-54;  
 Matches 262; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 238 gtcatctcccgctctatctctgttaacacatctcagaagaaggttcagatgtat 297  
 |||||

DB 1 gtcatctcccgctctatctctgttaacacatctcagaagaaggttcagatgtat 60  
 |||||

OY 298 gggctgatacgaagcctagaagaataagacaaatattgtgaagcacataaagt 357  
 |||||  
 DB 61 gggctgatacgaagcctagaagaataagacaaatattgtgaagcacataaagt 120  
 |||||

OY 358 ttcatcaatccatccagcagagatgagacatttgagacaggttccgccaagacgtcacca 417  
 |||||  
 DB 121 ttatcaantccatccagcagagatgagacatttgagacaggttccgccaagacgtcacca 180  
 OY 418 agtgccttccctaggtattatcttcacatccactttgtccaactaccgtctctgc 477  
 |||||  
 DB 181 agtgccttccctaggtattatcttcacatccactttgtccaantaccgtcttctgc 240  
 OY 478 taatgtacctgttcccgagcaactatgatcag 511  
 |||||  
 DB 241 taatgtacctgttcccgagcaactatgatcag 274

## RESULT 11

AAH30637  
 ID AAH30637 standard; cDNA; 398 BP.

AC AAH30637;

DT 27-JUL-2001 (first entry)

DE Human colon cancer cell line Km12L4-A cDNA library derived sequence #571.

XX Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;  
 detection; colon cancer cell line Km12L4-A; ss.

XX Homo sapiens.

PN WO200018916-A2.

PD 06-APR-2000.

PF 23-SEP-1999; 99WO-US222226.

PR 28-SEP-1998; 98US-0102161.

PR 28-SEP-1998; 98US-0102180.

PR 29-SEP-1998; 98US-0102380.

PR 08-OCT-1998; 98US-0103815.

PR 27-OCT-1998; 98US-0105877.

PA (CHIR ) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
 PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;  
 PI Lamson G, Drmanac R, Cirkenjakov R, Dickson M, Drmanac S, Labat I;  
 PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;

DR WPI; 2000-293155/25.

PT Polynucleotide library comprising 1079 defined sequences, useful in  
 the form of an array to detect cancer or susceptibility to cancer -  
 PS Claim 1; Page 347; 502pp; English.

XX The present invention describes a library of polynucleotides comprising  
 CC 1079 nucleotide sequences (given in AAH30637 to AAH31145). Also described  
 CC are: (1) an isolated polynucleotide (I) having at least 90% identity to  
 CC one of the 1079 sequences; (2) a recombinant host cell containing (1);  
 CC (3) an isolated polypeptide (II) encoded by (1); (4) an antibody that  
 CC specifically binds to (II); (5) a vector comprising (I); and (6) a method  
 CC of detecting differentially expressed genes correlated with a cancerous  
 CC state of a mammalian cell comprising detecting a gene product encoded by  
 CC 65 of the 1079 sequences given in the specification. The polynucleotides  
 CC are used to monitor patients having (or susceptible) to cancer to detect  
 CC potentially malignant events at a molecular level before they are  
 CC detectable at a gross morphological level. The polynucleotides are also  
 CC useful for monitoring the efficacy of various therapies and preventive  
 CC interventions. Polynucleotide probes based on the disclosed sequences  
 CC are useful for chromosome mapping and detection of transcription levels.  
 CC The 1079 polynucleotide sequences were derived from a human colon cancer  
 CC cell line Km12L4-A cDNA library.





```

Qy 1798 tgtattacagccctctgtagctcaactctgctccctcactgtagtcagacgacgca 1857
Db 441 www. .... 500
Qy 1856 actgtaactgacagtgcctctctctgagcactgacacactgtaaggtactaatt 1917
Db 501 www. .... 560
Qy 1918 actgcccagccctgggagatcagagagagctgcactgtagttaggtgttagcttt 1977
Db 561 www. .... 620
Qy 1978 tgtgtgcatcagtgactgaactctgtaataactattglaaatgcatgaagcactg 2037
Db 621 www. .... 680
Qy 2038 ttcttaaacccaagtaagaactcttgaaacctgttgatgtaaaaaa 2097
Db 681 www. .... 740
Qy 2098 aaaaaaaaaaaaaaaaaa 2118
Db 741 www. .... 761

RESULT 15
AAF58257
ID AAF58257 standard; DNA; 936 BP.
XX
AC AAF58257;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1954.
XX
KM Electron-transfer group; ETM; mismatch; genotyping;
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PE 26-JUL-2000; 2000MO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PI hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

```

Query Match 4.1%; Score 87; DB 22; Length 936;

```

Best Local Similarity 0.8%; Pred. No. 7.5e-12;
Matches 6; Conservative 435; Mismatches 300; Indels 0; Gaps 0;
Qy 1378 tttaacgtgatgtgctcagttcatgtctcaggaagcacttttcccaagtggaat 1437
Db 21 www. .... 80
Qy 1438 cctatttggcttggagcagtcacactcactcactgtaactacatctgtggtttt 1497
Db 81 www. .... 140
Qy 1498 tgcgttagaanaattttggtggtgtaaaaacagcactccttggctggagcactgtgtc 1557
Db 141 www. .... 200
Qy 1558 catgcatgactgtggtgttccctccatcccttctgatatgacaaataaagtgtt 1617
Db 201 www. .... 260
Qy 1618 ttgttttttgacactcactgcatggtgctaaccactctttttaaacccttgaaca 1677
Db 261 www. .... 320
Qy 1678 ccttttctgtagtgtaacttcagagatactctattgaaagataacagagatlaaca 1737
Db 321 www. .... 380
Qy 1738 gtgctcttgaaccctccctcaatglttctagccttcaactcactgtcttcttctgggc 1797
Db 381 www. .... 440
Qy 1798 tgtattacagccctctgtagctcaactctgctgcctcactgtagtcaagtcaca 1857
Db 441 www. .... 500
Qy 1858 actgtaactgacagtggtgctctctctgagcactgacacactgtaaggtactaatt 1917
Db 501 www. .... 560
Qy 1918 actgcccagccctgggagatcagagagagctgcactgtagttagttaggttt 1977
Db 561 www. .... 620
Qy 1978 tgtgtgcatcagtgacttagagttctgtaataactattgtaaatgcatgaagcactg 2037
Db 621 www. .... 680
Qy 2038 ttcttaaacccaagtaagaactcttgaaacctgttgatgtaaaaaa 2097
Db 681 www. .... 740
Qy 2098 aaaaaaaaaaaaaaaaaa 2118
Db 741 www. .... 761

```

Search completed: December 20, 2001, 18:54:40  
Job time: 5171 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2001, 10:10:52 ; Search time 13.29 Seconds

(without alignments)  
993.178 Million cell updates/sec

Title: US-09-868-474-2

Perfect score: 1912

Sequence: 1 MALSRVCMARSAMVGSATVP.....LSLLHNVLSTNYLSTR 360

Scoring table:

BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	171.5	9.0	1013	1 A60D_DROME	P91927 drosophila
2	92	4.8	656	1 SYM_HELPJ	Q92K99 helicobacte
3	91.5	4.8	1120	1 YB85_YEAST	P38329 saccharomyc
4	89	4.7	276	1 D101_ARATH	005211 arabidopsis
5	86	4.5	3066	1 POLG_BCMVN	065399 b genome po
6	85.5	4.5	2231	1 SENL_YEAST	000416 saccharomyc
7	84	4.4	561	1 SYR_CHLPP	Q92Y73 chlamydia p
8	84	4.4	650	1 SYM_HELPJ	P56127 helicobacte
9	83.5	4.4	603	1 UVRG_CHLPP	Q926W6 chlamydia p
10	83.5	4.4	738	1 NU5C_LIGVU	091143 ligustrum v
11	82.5	4.3	2199	1 DPOE_SCHPO	P87154 schizosach
12	81.5	4.3	816	1 AD15_RAT	094YV0 r adam 15 p
13	81.5	4.3	1230	1 PRPM_HUMAN	000411 homo sapien
14	81	4.2	1373	1 HRP1_SCHPO	Q90425 schizosach
15	81	4.2	1447	1 TOP2_DROME	P15348 drosophila
16	79.5	4.2	573	1 CYDC_ECOLI	P23886 escherichia
17	79.5	4.2	2215	1 MY7A_HUMAN	013402 homo sapien
18	79	4.1	500	1 Y039_BORBU	051068 borrelia bu
19	79	4.1	1181	1 HAIR_RAT	P97609 rattus norv
20	79	4.1	1182	1 HAIR_MOUSE	Q61645 mus musculu
21	78.5	4.1	469	1 P49_STR1I	P06108 streptomyce
22	78.5	4.1	505	1 PR12_MOUSE	P36100 mus musculu
23	78	4.1	536	1 CCA4_DROME	Q9VE00 drosophila
24	78	4.1	800	1 PR11_YEAST	P08468 saccharomyc
25	78	4.1	847	1 HEX_VIBVU	Q04786 vibrio vuln
26	77.5	4.1	310	1 HTPX_HELPJ	Q25582 helicobacte
27	77.5	4.1	383	1 HOXV_ALICU	P31914 alcaligenes
28	77.5	4.1	398	1 PARA_ECOLI	P07620 notophthal
29	77.5	4.1	505	1 RRG_NOTVI	P18516 notophthal
30	77.5	4.1	2747	1 FAF_DROME	P55824 drosophila
31	77	4.0	312	1 Y003_BPL2	P42538 bacteriopia
32	77	4.0	708	1 YB40_HUMAN	Q94100 homo sapien
33	77	4.0	873	1 COX1_ACACA	Q37370 acanthamoeb

34	76.5	4.0	206	1 ICT1_HUMAN	Q14197 homo sapien
35	76.5	4.0	346	1 NU2M_CORJA	P24971 coturnix co
36	76.5	4.0	436	1 UH32_HSVSA	Q01040 herpesvirus
37	76.5	4.0	500	1 PRG1_HUMAN	Q16647 homo sapien
38	76.5	4.0	722	1 GRE4_CHLPP	Q92794 chlamydia p
39	76.5	4.0	842	1 ABC6_HUMAN	Q9NP58 homo sapien
40	76.5	4.0	1878	1 BRCL_CANPA	Q95153 canis famli
41	76.5	4.0	2472	1 SPCN_HUMAN	Q13813 homo sapien
42	76	4.0	380	1 KPT3_HUMAN	Q07002 homo sapien
43	76	4.0	1067	1 STI_PYRAB	Q9V072 pyrococcus
44	75.5	3.9	310	1 HTPX_HELPJ	Q92K99 helicobacte
45	75.5	3.9	452	1 GP13_YEAST	P32363 saccharomyc

## ALIGNMENTS

RESULT 1  
ID A60D\_DROME STANDARD: PRT: 1013 AA.  
AC P91927; Q9W160;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DE 20-AUG-2001 (Rel. 40, Last annotation update)  
DE CALCIUM-BINDING MITOCHONDRIAL PROTEIN ANON-60DA.  
GN ANON-60DA OR CG4589  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abriil J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,  
RA Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu I., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brottler P.,  
RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.W., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Foubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ijegyem C.,  
RA Jaitani M.E., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshirei A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
RA Palenzuela M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reihartz K., Remington K., Saunders R.D.C., Scheefel F., Shen H.,  
RA Shue B.C., Stiden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
RA Splier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner C., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195(2000).

```

RN [2]
RP SEQUENCE OF 626-944 FROM N.A.
RC TISSUE-Ovary:
RX MEDLINE-99168769: Pubmed-10071211;
RA Caggese C., Ragone G., Perrini B., Moschetti R., de Pinto V.,
RA Calzati R., Barsanti P.;
RT "Identification of nuclear genes encoding mitochondrial proteins:
RT isolation of a collection of D. melanogaster cDNAs homologous to
RT sequences in the Human Gene Index database.";
RL Mol. Gen. Genet. 261:64-70(1999).
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
CC FRAMESHIFTS IN POSITIONS 920 AND 930.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AEO03464; AAF47217.1; -
DR EMBL: Y10912; CAAT1853.1; ALT_FRAME.
DR Flybase: FBgn0019886; CG4598.
DR InterPro: IPR001813; 60s.ribosomal.
DR InterPro: IPR002024; Bacterioferritin.
DR InterPro: IPR003871; DUF223.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000861; REM_repeat.
DR InterPro: IPR002555; REM.
DR InterPro: IPR001236; 1dh.
DR Pfam: PF00428; 60s.ribosomal; 1.
DR Pfam: PF01334; Bacteriofer; 1.
DR Pfam: PF02721; DUF223; 1.
DR Pfam: PF00036; ehand; 2.
DR Pfam: PF02185; HRI; 1.
DR Pfam: PF00056; 1dh; 1.
DR Pfam: PF01605; RE1; 1.
DR SMART: SMO0054; Eph; 2.
DR PROSITE: PS00018; EF_HAND; 2.
DR Mitochondrion: Calcium-binding; Repeat.
DR CA_BIND 700 711 EF_HAND 1 (POTENTIAL).
DR CA_BIND 783 794 EF_HAND 2 (POTENTIAL).
DR CONFLICT 687 687 E -> Q (IN REF. 2).
DR CONFLICT 690 690 K -> G (IN REF. 2).
DR CONFLICT 717 717 Q -> H (IN REF. 2).
DR CONFLICT 736 736 E -> D (IN REF. 2).
DR CONFLICT 740 740 K -> R (IN REF. 2).
DR CONFLICT 761 761 K -> E (IN REF. 2).
DR CONFLICT 943 943 G -> S (IN REF. 2).
SQ SEQUENCE 1013 AA; 113579 MW; 8C710380263F262E CRC64;

```

Query Match 9.0%; Score 171.5; DB 1; Length 1013;  
 Best Local Similarity 21.7%; Pred. No. 2.1e-07;  
 Matches 66; Conservative 54; Mismatches 119; Indels 65; Gaps 8;

```

QY 100 AKKARRITNMKKINIKFHQ-----LPRMEHLRQFQDV 135
DQ 175 AKPKPRRTIRIMDELVHYHGFRLLEFDVAICSLKRLWRVNLGKTLTRERKQLORTTSDL 234
QY 136 TKCFELGISIPRANVLVFLMYLFP-----RQLLRHWTK-----QQ 176
DQ 235 FRLLPFSYFIVPMEELLPFLTFKFPQMLSTQSTDRQEKLOSLSVRLVAKFLQQ 294
QY 177 T-DELDIYHAKRQSHPEIISYLEKVIPLISDAGLRWLRLDCTKIORGTHPAIHD-ILA 234
DQ 295 TLDDQMPVGH---KHSSEAKQFEAF-----TKINPPEPVSNDILIK 335
QY 235 LRECFSNHPLGMNOLQALHVAKLSRAMLTLSYLPRLRHRKTKTHVTIHOQKALAKLG 294

```

```

DQ 336 FAKRPDDE-ITLDSLSREOLAACRYLEINTIGTTLLRFQRLKRSIANDDRIAREG 394
QY 295 IGLTAQEVKSAACYLGLNSTHIGEDRCRTMGWLOISCSLKEAELSLHLNVLLSTN 354
DQ 395 VBSLDLELQQAQKARGMAYGLTERLRFLQKEMIDLSINQVPTPTLLLSRTMLISDD 454
QY 355 YLGT 358
DQ 455 SITT 458

```

## RESULT 2

SYM\_HELPJ STANDARD; PRT; 656 AA.

AC 092KG9;

DT 30-MAY-2000 (rel. 39, Created)

DT 30-MAY-2000 (rel. 39, Last sequence update)

DT 20-AUG-2001 (rel. 40, Last annotation update)

DE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA LIGASE)

DE (METERS).

GN METG OR JHP0967.

OC Helicobacter pylori J99 (Campylobacter pylori J99).

OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;

OC Helicobacter.

OC NCBI\_TaxID=85963;

RP [1]

RP SEQUENCE FROM N.A.

RA MEDLINE-99120557: Pubmed-9923682;

RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Dolg P.C.,

RA Smith D.R., Noonan B., Guild B.C., deGange B.L., Carmel G.,

RA Tumundo P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,

RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Voyis G.F.,

RA Trust T.J.;

RT "Genomic sequence comparison of two unrelated isolates of the human

RT gastric pathogen Helicobacter pylori.";

RL Nature 397:176-180(1999).

CC -1- FUNCTION: IT IS PROBABLY ESSENTIAL FOR CELL SURVIVAL, BEING

CC REQUIRED NOT ONLY FOR ELONGATION OF PROTEIN SYNTHESIS BUT ALSO

CC FOR THE INITIATION OF ALL RNA TRANSLATION THROUGH INITIATOR

CC TRNA(EMET) AMINOACYLATION.

CC -1- CATALYTIC ACTIVITY: ATP + L-METHIONINE + TRNA(MET) = AMP +

CC PYROPHOSPHATE + L-METHIONYL-TRNA(MET).

CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.

CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.

CC STRONG. TO CYSTEINYL-TRNA SYNTHETASE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: AEO01525; AAD06543.1; -

DR InterPro: IPR002300; trna-synt\_1a.

DR InterPro: IPR001412; trna-synt\_1.

DR InterPro: IPR002304; trna-synt\_met.

DR InterPro: IPR002547; trna\_bind.

DR Pfam: PF00133; trna-synt\_1; 1.

DR PRINTS: PR01041; TRNASYNTMET.

DR PROSITE: PS00178; AA TRNA LIGASE I; FALSE NEG.

KW Aminoacyl-TRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

KW trna-binding; Complete proteome.

KW SITE 11 21 "HIGH" REGION.

FT SITE 301 305 "KMSKS" REGION.

FT BINDING 304 304 ATP (BY SIMILARITY).

FT DOMAIN 561 654 TRNA BINDING.

SQ SEQUENCE 656 AA; 75433 MW; 8F22F3BA1F9B9756 CRC64;

Query Match 4.8%; Score 92; DB 1; Length 656;  
 Best Local Similarity 17.7%; Pred. No. 1.5;  
 Matches 49; Conservative 51; Mismatches 113; Indels 64; Gaps 8;

DB 28 LQGRSLGACGARRSSLLHSFKADYKNLMSYVYTKAIKNGYHRLGHPREFYLYT 87  
 204 LSTRISFEWGIDLPKKMN-DPKHYVYVWLDALNYSALG----- 243

DB 88 IFMGLIOMLADAKKARRIKTNMKNHFKHOLPYREMEHLRQFDYTKCLFLGIISTP 147  
 244 -YINGLONKMAHFERAHI---VGKILRFHAIYWP-----AFIMSLINLP 284

DB 148 PFANYVELLMYLEPRLRLRHFWT-----PKQOTDFDIYHAFKROSHPEIISYLEKY 201  
 285 LF-----KOLCVHGMWTEGVMKSLGNVDAQMLAMEYGEELRYFLLE 331

DB 202 IPLISDAGLRMR--LNDLCIKIQRGHPAIDHILARECSNHPIDGMNOLQALHYKALS 259  
 332 VPPGQDDESKKALVERINANNLNDLGNLRLGMAKKYFNSLKSTKITAYPRELER 391

DB 260 AMLL---TSYLPPLRLRKTHTTVIHQDLKALAK 292  
 392 AHQIDNANSPFKMOLHKALELPNIYDFLNKLAK 428

RESULT 3  
 YB85\_YEAST  
 ID YB85\_YEAST STANDARD; PRT: 1120 AA.  
 AC P38329;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE HYPOHETICL 124.0 KDA PROTEIN IN PBP2-ABD1 INTERGENIC REGION.  
 GN YBR235W OR YBR1601.  
 OS Saccharomyces cerevisiae (Baker's yeast)  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN RA SEQUENCE OF 1-975 FROM N.A.  
 RC STRAIN=S288C;  
 RA Dubois E., el Bakhouy M., Glansdorff N., Messenguy F., Plerard A.,  
 Scherens B., Vierendeels F.;  
 RL Submitted (Aug-1994) to the EMBL/Genbank/DBJ databases.  
 RP [2]  
 RA SEQUENCE OF 596-1120 FROM N.A.  
 RC STRAIN=S288C;  
 RA Aljinovic G., Pohl F.M., Pohl T.M.;  
 RL Submitted (Aug-1994) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC -1- SIMILARITY: TO C.ELEGANS K02A2.3.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: Z66104; CA85198.1; -  
 DR PIR: S46111; S46111.  
 DR SGD: S0000439; YBR235W.  
 DR InterPro: IPR002293; AA\_rel\_permease\_1.  
 DR InterPro: IPR002027; Amino\_acid\_permease.  
 DR Pfam: PR00324; aa\_permeases\_1.  
 KW Hypothetical protein; Transmembrane.  
 FT TRANSMEM 64 84 POTENTIAL.  
 FT TRANSMEM 86 106 POTENTIAL.  
 FT TRANSMEM 146 166 POTENTIAL.  
 FT TRANSMEM 194 214 POTENTIAL.  
 FT TRANSMEM 222 242 POTENTIAL.  
 FT TRANSMEM 284 304 POTENTIAL.

FT TRANSMEM 321 341 POTENTIAL.  
 FT TRANSMEM 361 381 POTENTIAL.  
 FT TRANSMEM 394 414 POTENTIAL.  
 FT TRANSMEM 431 451 POTENTIAL.  
 FT TRANSMEM 482 502 POTENTIAL.  
 FT TRANSMEM 598 618 POTENTIAL.  
 FT TRANSMEM 620 640 POTENTIAL.  
 FT TRANSMEM 755 775 POTENTIAL.  
 SQ SEQUENCE 1120 AA; 123998 MW; 299C99C0C8C35ECC CRC64;

Query Match 4.8%; Score 91.5; DB 1; Length 1120;  
 Best Local Similarity 19.5%; Pred. No. 3.2; Indels 113; Gaps 16;  
 Matches 68; Conservative 48; Mismatches 120;

DB 78 HPPREFIYLTIFPMGLIOMLADAKKARRIKTNMKNHFKHOLPYREMEHLRQFDYTK 137  
 403 HSPLSYLIFFWTILQT-CLFSDVNKIATFT-----MTPLMTFVVMNLA 445

DB 138 CLFLGIISTPFP-----ANYLVFLMLYLPRL 165  
 446 CFLLISSAPNRPSPFKYFNRYTTAIGALLSVAMLIYDGISASVLFAMILLFLE---- 501

DB 166 LIRHFWPKQOTDFLD--IYHAFK---QSHPEIISYLE-KVIPLEISDAGLRMRITDICT 219  
 502 -IHYFSPKSGDVOSQSLIYHOVKYLLRLNODIKIKRPQILLFVDPRTSMNLIRCN 560

DB 220 KIQRG-----THPAIDHILALRECSNHPIDGMNOLA-----LHYKALSRAMLLYSYL 267  
 561 HIKKGLYLIGHAV-----TADPKQINELKQOKAMMKIRDMAAIKAFQVGTG 611

DB 268 PPLRLRHR-----LKHHTYIHLQDLKALAKLGIGQLTAQ-----EVR-----S 305  
 612 PSLINGIRNVPFGSLGCKRPNTIVAGFFDLBSYKHHIPQSRSONNLOKQVEIKATVPRS 671

DB 306 ACYLGLNSTHAGEDRCRTW---LGEMLQISCSLKEALELLLNHNVL 350  
 672 TCSQVKIN-VPLPTECKNEKRVNQGVQVQI-----VEDLSLMSNIAI 714

RESULT 4  
 D101\_ARATH  
 ID D101\_ARATH STANDARD; PRT: 276 AA.  
 AC 005211;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE DNA-DAMAGE-REPAIR/TOLERATION PROTEIN DRT101 PRECURSOR.  
 GN DRT101.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eustoids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RC STRAIN=CV COLIMBIA;  
 RA MEDLINE-93320380; PubMed-8329681;  
 RA Pang Q., Hays J.B., Rajagopal I., Schaefer T.S.;  
 RT "Selection of Arabidopsis cDNAs that partially correct phenotypes of  
 RT Escherichia coli DNA-damage-sensitive mutants and analysis of two  
 RT plant cDNAs that appear to express UV-specific dark repair  
 RT activities";  
 RL Plant Mol. Biol. 22:411-426(1993).  
 CC -1- SIMILARITY: STRONG, TO THE C-TERMINAL HALF OF YEAST  
 CC PHOSPHOACETYLGLUCOSAMINE MUTASE (PCM1).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>



QY 294 GIGGL-THOEVKASCYLGNLSTHGEDRCFTWLGEMLOISCSLKEALSLLLNNVLLS 352  
 Db 970 --GRKSSRRKFFVSACFMNA--QTHLGNAR-----ITISKNVEV 1004  
 QY 353 TNYLSTR 360  
 Db 1005 TN-LGVR 1011

RESULT 6  
 SENL\_YEAST STANDARD: PRT: 2231 AA.  
 AC 000416; Q06448;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE TRNA-SPlicing ENDONUCLEASE POSITIVE EFFECTOR.  
 GN SEN1 OR YIR430W OR L9576.1.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE OF 120-2231 FROM N.A.  
 RX MEDLINE=92236590; PubMed=1569945;  
 RA Demarini D.J., Winey M., Ursic D., Webb F., Culbertson M.R.;  
 RT "SEN1, a positive effector of trna-splicing endonuclease in  
 RT Saccharomyces cerevisiae."  
 RL Mol. Cell. Biol. 12:2154-2164(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
 RA Fajello A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,  
 RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,  
 RA Johnston L., Langston T., Latreille P., Le T., Mardis E., Meneses S.,  
 RA Miller N., Nan M., Pauley A., Peluso D., Rifkin L., Riles L.,  
 RA Taich A., Trevasaki E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,  
 RA Wilson R., Waterston R.;  
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
 CC -I- FUNCTION: REQUIRED FOR ENDONUCLYTIC CLEAVAGE OF INTRONS FROM  
 CC ALL FAMILIES OF PRECURSOR TRNAS. MAY BE ONE OF SEVERAL COMPONENTS  
 CC OF A NUCLEAR-LOCALIZED SPLICING COMPLEX. SEN1 IS ESSENTIAL FOR  
 CC VEGETATIVE GROWTH. ESSENTIAL FUNCTIONS OF THE PROTEIN RESIDE IN  
 CC THE C-TERMINAL 1214 AMINO ACIDS.  
 CC -I- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -I- SIMILARITY: BELONGS TO THE DNA2/NAM7 HELICASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC DR EMBL; M74589; AAB63976.1;  
 CC DR EMBL; U20939; AAB67502.1;  
 CC DR PIR: A4387; A44387.  
 CC DR SGP: S0004422; SEN1.  
 CC DR LRNA Processing; SEN1.  
 CC KM LRNA Processing; ATP-binding; Hydrolase; Nuclease.  
 CC FT NP\_BIND 1357 1364 ATP (POTENTIAL).  
 CC FT DOMAIN 1908 1961 LYS-RICH (BASIC).  
 CC FT DOMAIN 1909 1927 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 CC FT VARIANT 1747 1747 G -> A (IN SEN1-1; GIVES RISE TO A  
 CC TEMPERATURE SENSITIVE MUTANT).  
 CC FT SEQUENCE 2231 AA; 252495 MW; 88F0FF95B3A8BD89 CRC64;

Query Match 4.5%; Score 85.5; DB 1; Length 2231;  
 Best Local Similarity 18.2%; Pred. No. 26;  
 Matches 66; Conservative 63; Mismatches 140; Indels 93; Gaps 15;

QY 40 PRSSKHLSPKADYKINMSYVVTKTAINCKYHREPLGRHFPRIYLTTPMKGLQMLMAD 99  
 Db 193 PHMLRLKKOLKATFEALFEKFFY-DTK-----HRLDVTNP--LSIKFISGVCWCE 242  
 QY 100 AKKARRIKITNMWK-----HNKIFHQLPYREHNLROFQODYTKCLGIIISIPFANV 152  
 Db 243 GSK-----EENEMSKRAFLKLDYSRNF-----INLSNLPDIEEYVTHILFLQDPANW 291  
 QY 153 LVFL-----LMYLFPROLLIRHFWTPKO-----QTFEDIYHARRKOS 190  
 Db 292 TELVVSQFWSRLPLPVNLPFKDWEIEFYQPKVNESLKTFFKPLPEIFPMWNNHLSKV 351  
 QY 191 HPEITYLEKVIPLDISDAGLRWL---TDLCTKIQGTGTHPAIDILALRECF----- 239  
 Db 352 HDKPLDFLLR-----GLTFELNKGESEFWSKIEPFTSHLIDILFNDSPIKILKI 403  
 QY 240 SNHPLGNNQOLALHVKALSRAMLITSLPPLLRHRLKTTTV--IHQDLKALAKIGIG 297  
 Db 404 QDNPIVEHOTEVYFQLTGVSVDLSWTLF--FYHALSPSKRIQWVKVMAFLRIITANY 460  
 QY 298 LTAQEVKASCYLGNLSTHGEDRCFTWLGEMLOISCSLKEALSLLLNNVLLS 352  
 Db 461 PSLSKTPKACILMSATL-----LRAVLTIKENRAMLYKNDEFEVLLTK 506  
 QY 353 TN 354  
 Db 507 TD 508

RESULT 7  
 SYR\_CHLPN STANDARD: PRT: 561 AA.  
 ID SYR\_CHLPN  
 AC 0927Y3; Q9JG78;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ARGINYL-TRNA SYNTHETASE (EC 6.1.1.19) (ARGININE--TRNA LIGASE) (ARGRS).  
 GN ARGS OR CPN0570 OR CP0179.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CWL029;  
 RA MEDLINE=99206606; PubMed=10192388;  
 RX Kalmn S., Mitchell W., Marathe R., Tamme C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heideberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Uetzerback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Eisen M., Nelson W., Deboy R., Koloney J., McClarty G., Salzberg S.L.,  
 RA Gwin J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
 RT pneumoniae AR39."  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kihara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CWL029 from USA."  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 CC -I- CATALYTIC ACTIVITY: ATP + L-ARGININE + TRNA(ARG) = AMP +  
 CC PYROPHOSPHATE + L-ARGINYL-TRNA(ARG).  
 CC -I- SUBUNIT: MONOMER (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE001641; AAD18710.1; -  
 DR EMBL: AE002179; AAF38053.1; -  
 DR EMBL: AP002547; BAA98776.1; -  
 DR TIGR: CP0179; -  
 DR InterPro: IPR001278; trna-synt\_1d.  
 DR InterPro: IPR001412; trna-synt\_1t.  
 DR Pfam: PF00750; trna-synt\_1d; 1.  
 DR PRINTS: PR01038; TRNASYNTHARG.  
 DR PROSITE: PS00178; AA\_TRNA\_LIGASE\_1; 1.  
 KM Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KM Complete proteome. 133 "HIGH" REGION.  
 FT SITE 125 369 "KMSKS" REGION.  
 FT BINDING 369 370 App (BY SIMILARITY).  
 FT SEQUENCE 561 AA; 63368 MW; 6386F0FF190C22DB CRC64;

Query Match 4.4%; Score 84; DB 1; Length 561;  
 Best Local Similarity 23.7%; Pred. No. 6.5;  
 Matches 44; Conservative 23; Mismatches 59; Indels 60; Gaps 8;

QY 178 DFDDIYHAFKQSH-----PEIISYLEK-----VIPLI---SDAGLR 211  
 DB 242 DIIDIVVEKGESEYNFLEPIEIDLEKGLLVNSNDKACVFHEARSIPPMVQSGGVY 301  
 QY 212 WRLDLCIKQRTGTHPAIDILALRECFSNHPLGMNOLAHVKAISRAMLTSTYLPPL 271  
 DB 302 YATDILAMRYIEDHAKIIVTD-----LG---GSLHQLLEAVTAIAAGYQPGI 351  
 QY 272 LRH-----RLKTHP-----TVIHQLDLAKLIGIGTAAQEVKSAC 308  
 DB 352 FSHVGFELVDPQCKLTKRSNGENKRLLELDTAIEKAELAKE-HRPELDEAIDERPAP 410  
 QY 309 LRGIN 314  
 DB 411 VIGINA 416  
 RESULT 8  
 SYM\_HELPY STANDARD; PRT; 650 AA.  
 AC P56127;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA LIGASE)  
 DE (METS).  
 OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 OC Helicobacter.  
 OX NCBI\_TaxID=210;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=26695 / ATCC 700392;  
 RX MEDLINE=9739467; Pubmed=9252185;  
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,  
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,  
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,  
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,  
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,

RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,  
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the gastric pathogen Helicobacter  
 RT pylori".  
 RL Nature 388:539-547(1997).  
 CC -1- FUNCTION: IT IS PROBABLY ESSENTIAL FOR CELL SURVIVAL, BEING  
 CC REQUIRED NOT ONLY FOR ELONGATION OF PROTEIN SYNTHESIS BUT ALSO  
 CC FOR THE INITIATION OF ALL MRNA TRANSLATION THROUGH INITIATOR  
 CC TRNA(EMET) AMINOACYLATION.  
 CC -1- CATALYTIC ACTIVITY: ATP + L-METHIONINE + TRNA(MET) = AMP +  
 CC PYROPHOSPHATE + L-METHIONYL-TRNA(MET).  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.  
 CC STRONG: TO CSTEINYL-TRNA SYNTHETASE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE000557; AAD07483.1; -  
 DR TIGR: HP0417; -  
 DR InterPro: IPR002300; trna-synt\_1a.  
 DR InterPro: IPR001412; trna-synt\_1t.  
 DR InterPro: IPR002304; trna-synt\_mel.  
 DR InterPro: IPR002547; trna\_bind.  
 DR Pfam: PF00133; trna-synt\_1; 1.  
 DR Pfam: PF01588; trna\_bind; 1.  
 DR PRINTS: PR01041; TRNASYNTHET.  
 DR PROSITE: PS00178; AA\_TRNA\_LIGASE\_1; FALSE\_NEG.  
 KM Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KM trna-binding; Complete proteome.  
 FT SITE 9 21 "HIGH" REGION.  
 FT SITE 301 305 "KMSKS" REGION.  
 FT BINDING 304 304 ATP (BY SIMILARITY).  
 FT DOMAIN 555 648 TRNA BINDING.  
 FT SEQUENCE 650 AA; 74506 MW; CERPDA57B9F75DD3 CRC64;

Query Match 4.4%; Score 84; DB 1; Length 650;  
 Best Local Similarity 17.7%; Pred. No. 7.7;  
 Matches 49; Conservative 48; Mismatches 116; Indels 64; Gaps 8;

QY 28 IQLGRSGLAWAPRSSKTLHSFKADYKNLMYSVYTKTAINGKYHRLGRHPRFYIYT 87  
 DB 204 LSIPTSEFWGIPLPKKN-DEKHHVYVWLDLALNYASALG----- 243  
 QY 88 IMKGIQMLMADAKKARKRTMMKNIKFQDLPYREHNLROFRODYTKCLFGIISIP 147  
 DB 244 -YLNDIDNKMAHEECARH---VGKDLIRFAIYWP-----AFLMSLNP 284  
 QY 148 PRANTLVFLMYLPROFLIRHFWT-----PKQOTDFLIYHAFRKOSHPEIISYLEKV 201  
 DB 285 LF-----KOLCVHGMWTIEGVKMSKSGVNLQDKAMETGIELFLPFLRE 331  
 QY 202 IPLISDAGLRW--LTDLCTKIQRTGTHPAIDILALRECFSNHPLGMNOLAHVKAISR 259  
 DB 332 VPGQDGFSEKALIERINANLNLDGLMLNRLGKAMARKYFNHSIKSTIYAYYSKELEK 391  
 QY 260 AMLL-----TSTYLPPLLRHRLKTHHTVYHODKALAK 292  
 DB 392 VHOILDNANSFVKMQLAKALEELFNVDPLNKLIAK 428

RESULT 9  
 UVRG\_CHLPPN STANDARD; PRT; 603 AA.  
 ID UVRG\_CHLPPN  
 AC Q926W6; Q92QA6;

DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE EXCINUCLEASE ABC SUBUNIT C.  
 GN UVRC OR CPN0940 OR CP0921.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CWL029;  
 RX MEDLINE=9206606; PubMed=10192388;  
 RA Kaiman S., Mitchell W., Marathe R., Lamme C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Ueberback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gaim M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Elsen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
 pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shira M., Hiraoka H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 from Japan and CML029 from USA.";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 CC -1- FUNCTION: THE ABC EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS  
 CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS  
 PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). ATTACHES TO THE  
 UVRA-UVRB COMPLEX, DISPLACING UVRA, AND THE DAMAGED DNA STRAND IS  
 NICKED ON BOTH SIDES OF THE DAMAGED SITE (BY SIMILARITY).  
 CC -1- SUBUNIT: CONSISTS OF THREE SUBUNITS: UVRA, UVRB AND UVRC.  
 CC -1- SUBCELLULAR LOCATION: CYTOSOL/PLASMA (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE UVRC FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: AE001673; AAD19078.1; -  
 DR EMBL: AE002251; AAF38706.1; -  
 DR EMBL: AP002548; BAA99148.1; -  
 DR TIGR: CP0921; -  
 DR InterPro: IPR001943; UVR.  
 DR InterPro: IPR001162; UVR\_C\_2.  
 DR InterPro: IPR000305; UVR\_C\_1.  
 DR Pfam: PF01541; Excl\_endo\_N\_1.  
 DR Pfam: PF02151; UVR\_1.  
 DR Pfam: PF005870; UVR\_C\_2; 1.  
 DR ProDom: PD005870; UVR\_C\_2; 1.  
 DR SMART: SM00465; G1YC; 1.  
 DR SOS response; Excision nuclease; DNA repair; Complete proteome.  
 KW SEQUENCE 603 AA; 69426 MW; 28017E4FA19FFC84 CRC64;  
 SQ

QY 88 IFMG-LQMLNADAKKARIRKTNMKINIKFHP--YREMEHRO--FRQDYTKCLFLG 142  
 Db 198 LFLKKEEYVKDLEKVIQKASD---NLEFQANRYRTISLKKQMAKQVKEKPHQN 253  
 QY 143 IISIPFANYIVFLMYLPPRO--LLIRHWPQKQOTDFLDIYHAERKQ---SHP--- 192  
 Db 254 IDALGLYHNKQVFTLLTVRSGLKLGARHFSFENAOEDDLSSFTLQYVVSQPIYK 313  
 QY 193 ELSILEKVIPLIS-----DAGIRMLVTLD-----CTKIQGTIHP-- 227  
 Db 314 ELTPLEPFLPSYLVNAESPRLRSKPTGKELDLAVRNKAVATTLPSSTLPYQ 373  
 QY 228 ALHDLALR-----ECFSN-----HPLGM-----NOLQ 250  
 Db 374 DQONILRMQYRYRECDNAINMGAHATGYIVFENNNGFQPKQYRTSIDSEKTONDLA 433  
 QY 251 ALHVALSRAMLTSLPPLLRHRLKTHYVYHQLKALAKLGIGULTADEVKSACTLR 310  
 Db 434 LLEEVLLRFHSLTALPDMIVDGGKTHYKTKKIOTLNLGIQVYIAKESN-HSR 492  
 QY 311 GLNSTHIGEDRCRTW 325  
 Db 493 GLNKERT--FCETP 504

RESULT 10  
 NUCS\_LIGVU  
 ID NUCS\_LIGVU STANDARD; PRT; 738 AA.  
 AC Q97LA3;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE NADH-PLASTOQUINONE OXIDOREDUCTASE CHAIN 5, CHLOROPLAST (EC 1.6.5.3).  
 GN NDHF.  
 OS Ligustrum vulgare (Common privet).  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Lamiales; Oleaceae; Ligustrum.  
 OX NCBI\_TaxID=13597;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kim K.-J., Jansen R.K., Olmstead R.C.;  
 RT "Multiple origins of sympetaly and associated floral characters.";  
 RL Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: NADH + PLASTOQUINONE = NAD(+) + PLASTOQUINOL.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC  
 DR EMBL: AF130164; AAF08126.1; -  
 DR InterPro: IPR001750; Oxidored\_q1.  
 DR InterPro: IPR002128; Oxidored\_q1\_C.  
 DR InterPro: IPR001516; Oxidored\_q1\_N.  
 DR Pfam: PF00361; Oxidored\_q1; 1.  
 DR Pfam: PF01010; Oxidored\_q1\_C; 1.  
 DR Pfam: PF00662; Oxidored\_q1\_N; 1.  
 DR Oxidoreductase; NAD; Plastocyanine; Chloroplast.  
 KW SEQUENCE 738 AA; 83477 MW; CD7BECABDC1BCAF CRC64;  
 SQ

Query Match 4.4%; Score 83.5; DB 1; Length 738;  
 Best Local Similarity 22.3%; Pred. No. 10;  
 Matches 39; Conservative 24; Mismatches 61; Indels 51; Gaps 7;

QY 59 YVVTKATNGKYNHFLGRRHPPFYLYITIPKGLQMLM--ADAKKA---RIKTKMMK 112  
 Db 450 YLTPFEGHLNHYFQVNSKQMTPFYSI-----SLWCKGDSKRIKNRFLKMNNSK 500

```

QY 113 HNIKQHPYREMEHLRQ-----FRQDYTK-----CLFLG 142
DB 501 SSSFFSKPYRSGKNVNRGPELTIIVHFNQKSYSPRESQNTMLPPLVLGILFLFWS 560
QY 143 IISPPRANVLVFLMLFPROLLIRHEFTWPKQOTDFLDIYHAFKROSHPELISY 197
DB 561 SLGI-FPNOELDLITLWPLPSIMLHOKW-----NDSIDWYEFKWDASFSVSIAV 609

RESULT 11
DPOE_SCHPO STANDARD: PRT: 2199 AA.
AC P87154;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA POLYMERASE EPSILON, CATALYTIC SUBUNIT A (EC 2.7.7.7) (DNA
POLYMERASE II SUBUNIT A).
GN CDC20 OR SPBC25H2.13C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=972;
RA Duroso G., Lye G., Bowman S., Church C., Wood V., Barrell B.G.,
RA Rajendream M.A., Connor R.E.;
RA Submitted (May-1997) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: DNA POLYMERASE II PARTICIPATES IN CHROMOSOMAL DNA
REPLICATION (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
N PYROPHOSPHATE + DNA(N).
CC -1- SUBUNIT: CONSISTS OF FIVE SUBUNITS (200 KDA, 80 KDA, 34 KDA, 30
KDA, AND 29 KDA) (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- DOMAIN: THE DNA POLYMERASE ACTIVITY DOMAIN RESIDES IN THE
N-TERMINAL HALF OF THE PROTEIN, WHILE THE C-TERMINUS IS NECESSARY
FOR COMPLEXING SUBUNITS B AND C (BY SIMILARITY).
CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
DIFFERENT REACTIONS OF DNA SYNTHESIS.
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. HIGH
SIMILARITY WITH MAMMALIAN DNA POLYMERASE EPSILON.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z93397; CAB08772.1; -
DR InterPro; IPR002064; DNA_POL_B.
DR Pfam; PF00136; DNA_POL_B_2.
DR PRINTS; PR00106; DNAPOLB.
DR SMART; SM00486; POLBC; 1.
KW Transferrase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Zinc-finger; Nuclear protein.
FT ZN_FING 2069 2155 BY SIMILARITY.
FT SEQUENCE 2199 AA; 252885 MW; AB0A5D0865FEBC3E CMC64;

Query Match 4.38; Score 82.5; DB 1; Length 2199;
Best Local Similarity 26.7%; Pred. No. 47;
Matches 36; Conservative 14; Mismatches 46; Indels 39; Gaps 5;
QY 177 TDFDIDYHAFKROSHPELISYLEKVIPLISDAGLRMLTJLCKRIQNGTIPALHDLIALR 236
DB 462 SDAAVATFELMKVYHPFTFS-LCNIIPLNDEYVR-----KGTGTLCELTLLTVE 509

```

```

QY 237 EC-----FSNHPLGMNQALHVKALS-----RAMLLTSYLPPLLRHR 275
DB 510 ACTKNILLPKNHVDASOKFFDGHLLASEYVGVHVESLSEGVFRSDLPFTNMDP----- 564
QY 276 LKRTFTVTHQDLKAL 290
DB 565 -KVYEFLLQLDKAL 578

RESULT 12
AD15_RAT STANDARD: PRT: 816 AA.
ID AD15_RAT
AC Q9QYV0;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ADAM 15 PRECURSOR (EC 3.4.24.-) (A DISINTEGRIN AND METALLOPROTEINASE
DOMAIN 15) (METALLOPROTEINASE-LIKE, DISINTEGRIN-LIKE, AND CYSTEINE-
DE RICH PROTEIN 15) (MDC-15) (METALLOPROTEINASE RGD DISINTEGRIN PROTEIN)
DE (METARGIDIN) (CRIL-7).
GN ADAM15 OR MDC15.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Sciatic nerve;
RX MEDLINE=20556183; PubMed=11102971;
RA Bosse F., Petzold G., Greiner-Petter R., Pippl U., Gallen C.,
RA Mueller H.-W.;
RT "Cellular localization of the disintegrin CRIL-7/rMDC15 mRNA in rat
RT PNS and CNS and regulated expression in postnatal development and
RT after nerve injury.";
RL Gila 32:313-327(2000).
CC -1- FUNCTION: MAY BE INVOLVED IN CELL-SURFACE PROTEOLYSIS, CELL
ADHESION OR INTRACELLULAR PROTEIN MATURATION.
CC -1- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC -1- SUBUNIT: INTERACTS WITH INTEGRIN ALPHA-V-BETA3, ENDOPHILIN I AND
SORTING NEXIN 9. ENDOPHILIN I AND SORTING NEXIN 9 PREFERENTIALLY
BIND THE PRECURSOR BUT NOT THE PROCESSED FORM OF ADAM15,
SUGGESTING THAT THE INTERACTION OCCURS IN A SECRETORY PATHWAY
COMPARTMENT PRIOR TO THE MEDIAL GOLGI (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN BRAIN, SPINAL CORD,
SCIATIC NERVE AND LUNG. EXPRESSED AT LOWER LEVELS IN ALL OTHERS
TISSUES. IN THE PERIPHERAL NERVOUS SYSTEM, EXPRESSED PREDOMINANTLY
BY SCHWANN CELLS. IN THE CENTRAL NERVOUS SYSTEM, PREFERENTIALLY
EXPRESSED BY NEURONAL CELLS.
CC -1- INDUCTION: IN RESPONSE TO SCIATIC NERVE INJURY.
CC -1- DOMAIN: THE CYTOPLASMIC DOMAIN INTERACTS WITH ENDOPHILIN I AND
SORTING NEXIN 9 (BY SIMILARITY).
CC -1- DOMAIN: DESINTEGRIN DOMAIN BINDS TO INTEGRIN ALPHA-V-BETA3 (BY
SIMILARITY).
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FUZIN ENDOPEPTIDASE (BY
SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
METALLOPROTEINASE); ALSO KNOWN AS THE REPOUSIN SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 SH3-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ251198; CAB61762.1; -
DR HSP; P17494; 1KST.

```

```

DR InterPro: IPR000130; Zn_MTPetase.
DR InterPro: IPR000591; EGF-like.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR001762; Disintegrin.
DR InterPro: IPR002870; Pcp_M12B_propep.
DR Pfam: PF00200; disintegrin_1.
DR Pfam: PF01562; Pcp_M12B_propep_1.
DR Pfam: PF01421; Reprolysin_1.
DR ProDom: PD000664; Disintegrin_1.
DR SMART: SM00050; Disin; 1.
DR SMART: SM00181; EGF_1.
DR PROSITE: PS00215; ADAM_MERO; 1.
DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE: PS00214; DISINTEGRIN_2; 1.
DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PROSITE: PS00546; CYSTEINE_SWITCH; FALSE_NEG.
KM Hydrolase; Metalloprotease; zinc; Signal; Glycoprotein; Zymogen;
transmembrane; EGF-like domain; SH3-binding.
FT SIGNAL 1 17
FT PROPEP 18 208
FT CHAIN 209 816
FT DOMAIN 209 698
FT TRAMEM 699 719
FT DOMAIN 720 816
FT DOMAIN 209 416
FT DOMAIN 423 510
FT DOMAIN 511 658
FT DOMAIN 659 687
FT DOMAIN 700 713
FT DOMAIN 768 774
FT SITE 803 809
FT SITE 180 180
FT METAL 350 350
FT ACT_SITE 351 351
FT METAL 354 354
FT METAL 360 360
FT DISULFID 325 411
FT DISULFID 482 495
FT DISULFID 669 669
FT DISULFID 663 675
FT DISULFID 677 686
FT CARBOHYD 57 57
FT CARBOHYD 239 239
FT CARBOHYD 391 391
FT CARBOHYD 394 394
FT CARBOHYD 608 608
FT CARBOHYD 613 613
SQ SEQUENCE 816 AA; 88051 MW; B9D2CE023266FC27 CRC64;

```

```

Query Match 4.3%; Score 81.5; DB 1; Length 816;
Best Local Similarity 22.7%; Pred. No. 17;
Matches 59; Conservative 33; Mismatches 71; Indels 97; Gaps 16;

```

```

QY 124 EMEHLROFRODY--TKCLFLGIIIS-----IPFANVL-----VFLMLFPPQLIR 168
DB 199 EQHNLRRLKRDVYETKIVELYADVNSEVKYRPFQQLNKRTLEALLDTFFQPLNVR 258
QY 169 -----HEWTPKQOQDFLDIYHAFKQSHPEIISTYLEKVIPLISDAGLRWRLTDLCTRIQ 222
DB 259 VALVGLLEAWT---QRDLIEM-----SSNPAY-----LDNFLRWRRRTDLLPLRP 299
QY 223 RETHPAIDHI--LALRECFNSNPIGLMNQLOALHVKALSRAMLITISYLPPLLRH---LK 277
DB 300 -----HDSAOLVTVTSGPMWGM-----AIDNSICSPDSCGVNMDHSTSLIG 343
QY 278 THTTVIHOLDKALAKIGI-----GQLTAQEV---KSACYLRLGLNSTHIGEDRC 322
DB 344 VASSIAHELGHIS---LGLDHDSPGNSCPGPRAPAKSCIMEASTDFLPLGLNFSN-----C 395
QY 323 RTW-----LGEW 329

```

```

DB 396 SRMALEKALDGMSCLEFW 415
RESULT 13
ID RPOK_HUMAN STANDARD; PRT; 1230 AA.
AC 000411; 060370;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE, MITOCHONDRIAL PRECURSOR (EC 2.7.7.6)
DE (MTRPOL).
GN POLRMT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-liver:
RX MEDLINE-97252399; PubMed-9097968;
RA Tiranti V., Savola A., Forti F., D'Apolito M.F., Centra M., Rocchi M.,
RA Zeviani M.;
RT *Identification of the gene encoding the human mitochondrial RNA
RT polymerase (h-mRPol) by cyberscreening of the Expressed Sequence
RL Hum. Mol. Genet. 6:615-625(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Barnes J., Danganan L., Poundstone P.,
RA Christensen M., Georgescu A., Avila J., Liu S., Altix C., Andreise T.,
RA Trankheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,
RA Bruce R., Thomas P., Quan G., Krommiller B., Arellano A.,
RA Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.O.,
RA Carraro A.V.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OR DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC RNA(N).
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -1- SIMILARITY: BELONGS TO THE PHAGE AND MITOCHONDRIAL RNA POLYMERASES
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U75370; AAB58255.1; -.
DR EMBL: AC004449; AAC06147.1; -.
DR MIM: 601778; -.
DR InterPro: IPR002092; RNA_pol_phage.
DR Pfam: PF00940; RNA_pol_1.
DR PROSITE: PS00900; RNA_POL_PHAGE_1; 1.
DR PROSITE: PS00489; RNA_POL_PHAGE_2; 1.
KW Transferase; DNA-directed RNA polymerase; Transcription;
KW Mitochondrion; Transist peptide.
FT TRANSIT 1 41
FT CHAIN 42 1230
FT ACT_SITE 922 922
FT ACT_SITE 991 991
FT ACT_SITE 1151 1151
FT CONFLICT 399 399
FT CONFLICT 983 983
SQ SEQUENCE 1230 AA; 138684 MW; 412E124D517FEC1A CRC64;

```



RL J. Mol.Biol. 205:1-13(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A..  
 RC STRAIN-BERKELEY;  
 RX MEDLINE-20196006; PubMed-10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Adamaitides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA Gordon R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champae M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abul J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
 RA Botkin D., Botchan M.R., Bouck T., Brokstein P., Brotlier P.,  
 RA Butkus K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dubler K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,  
 RA Fouts D.C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,  
 RA Jatali M., Kalusz F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Liao X., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon B., Nusskern D.R., Pacler J.M.,  
 RA Palazzolo M., Plutman G.S., Pan S., Pollard J., Puti V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Sidenklamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Wolley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhu Q., Zheng L.,  
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "the genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT  
 CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II  
 CC MAKES DOUBLE-STRAND BREAKS.  
 CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING  
 CC OF DOUBLE-STRANDED DNA.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH  
 CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES  
 CC RELAX ONLY NEGATIVE SUPERCOILS.  
 CC -1- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
 CC The European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X61209; CAA43523.1; -;  
 DR EMBL: AE003663; AAF53802.1; ALT\_SEQ.  
 DR PIR: S02160; S02160.  
 DR HSSP: P06786; 1BGW.  
 DR FLYBase: FBgn0003732; Top2.  
 DR InterPro: IPR000947; CBFA\_NFYB.  
 DR InterPro: IPR001241; DNA\_topoisoli.  
 DR InterPro: IPR003205; DNA\_topoisoliV.  
 DR InterPro: IPR003594; HATPase.C.  
 DR Pfam: PF00204; DNA\_topoisoli; 1.



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2001, 10:09:07 ; Search time 12.62 Seconds  
(without alignments)  
641.932 Million cell updates/sec

Title: US-09-868-474-2

Perfect score: 1912  
Sequence: 1 MALSRVCMARSAVMSAVTP.....LSLLHNVLSTNYIGTRR 360

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	124	6.5	586	2	US-08-630-822A-70
2	124	6.5	586	2	US-09-005-069-70
3	94	4.9	428	4	US-09-139-064-2
4	94	4.9	428	4	US-09-139-064-4
5	94	4.9	428	4	US-09-487-370-2
6	94	4.9	428	4	US-09-487-370-4
7	87	4.6	648	1	US-08-451-715A-4
8	84	4.4	1257	2	US-08-750-152A-2
9	80.5	4.2	147	3	US-08-946-329A-58
10	80	4.2	200	1	US-07-959-284-8
11	80	4.2	200	2	US-08-308-736A-8
12	80	4.2	200	5	PCT-US93-09649A-8
13	80	4.2	200	5	PCT-US93-09649-8
14	79	4.1	452	3	US-09-052-778-15
15	79	4.1	877	3	US-08-911-853-31
16	79	4.1	877	4	US-09-479-409-31
17	78.5	4.1	451	4	US-09-357-251-35
18	78.5	4.1	2629	2	US-08-751-189-4
19	78.5	4.1	2629	2	US-08-060-836-4
20	78.5	4.1	2629	4	US-09-184-445-4
21	77	4.0	402	1	US-08-036-210-15
22	77	4.0	402	2	US-08-449-609-15
23	77	4.0	898	2	US-08-036-210-22
24	77	4.0	898	2	US-08-449-609-22
25	76.5	4.0	500	2	US-08-578-709-15
26	76.5	4.0	1151	3	US-08-840-006-6
27	76.5	4.0	1200	3	US-08-840-006-5

28	75.5	3.9	472	4	US-08-976-255-17	Sequence 17, Appl
29	75.5	3.9	1005	1	US-08-089-986-3	Sequence 3, Appl
30	75.5	3.9	1005	1	US-08-478-585-3	Sequence 3, Appl
31	75.5	3.9	1005	1	US-08-717-312-3	Sequence 3, Appl
32	75.5	3.9	1005	2	US-08-266-408-3	Sequence 3, Appl
33	75.5	3.9	1005	5	PCT-US94-07886-3	Sequence 3, Appl
34	75	3.9	870	1	US-08-190-687B-25	Sequence 25, Appl
35	75	3.9	870	5	PCT-US94-00198-5	Sequence 5, Appl
36	75	3.9	1047	1	US-08-190-687B-8	Sequence 8, Appl
37	74.5	3.9	336	1	US-08-176-427B-6	Sequence 6, Appl
38	74.5	3.9	336	2	US-08-356-060A-10	Sequence 10, Appl
39	74.5	3.9	336	4	US-08-674-509B-10	Sequence 10, Appl
40	74.5	3.9	411	4	US-08-460-900C-10	Sequence 10, Appl
41	74.5	3.9	411	4	US-08-954-698-10	Sequence 4, Appl
42	74.5	3.9	492	1	US-07-783-705A-4	Sequence 4, Appl
43	74	3.9	326	2	US-09-055-097-6	Sequence 6, Appl
44	73.5	3.8	1064	3	US-08-726-214-8	Sequence 8, Appl
45	73	3.8	614	1	US-08-543-881-2	Sequence 2, Appl

#### ALIGNMENTS

RESULT 1  
US-08-630-822A-70  
; Sequence 70, Application US/08630822A  
; Patent No. 5840695  
; GENERAL INFORMATION:  
; APPLICANT: FRANK, GLENN R.  
; APPLICANT: HUNTER, SHIRLEY WU  
; APPLICANT: WALLENFELS, LYNDIA  
; TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS  
; TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS  
; NUMBER OF SEQUENCES: 107  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheridan Ross P.C.  
; STREET: 1700 Lincoln Street, suite 3500  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: U.S.A.  
; ZIP: 80203  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/630,822A  
; FILING DATE: 11-APR-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CONNELL, GARY J.  
; REGISTRATION NUMBER: 32,020  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 863-9700  
; TELEFAX: (303) 863-0223  
; INFORMATION FOR SEQ ID NO: 70:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 586 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: Xaa = any amino acid  
; LOCATION: 379  
; US-08-630-822A-70

Query Match 6.5%; Score 124; DB 2; Length 586;  
Best Local Similarity 27.0%; Pred. No. 3.3e-05;  
Matches 37; Conservative 29; Mismatches 69; Indels 2; Gaps 2;



APPLICANT: Biswas, Sanjoy  
 TITLE OF INVENTION: Response Regulator  
 FILE REFERENCE: GM10091  
 CURRENT APPLICATION NUMBER: US/09/139,064  
 CURRENT FILING DATE: 1998-08-25  
 EARLIER APPLICATION NUMBER: 60/060,714  
 EARLIER FILING DATE: 1997-09-09  
 NUMBER OF SEQ ID NOS: 9  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 4  
 LENGTH: 428  
 TYPE: PRF  
 ORGANISM: Streptococcus pneumoniae  
 US-09-139-064-4

Query Match 4.9%; Score 94; DB 4; Length 428;  
 Best Local Similarity 20.8%; Pred. No. 0.047;  
 Matches 68; Conservative 56; Mismatches 115; Indels 88; Gaps 16;

QY 29 QLRSGGLAWGAPRSSKHLSPKADVKNLMSYVYTKAKINCKYHREFLGRHPREYIYTI 88  
 DB 122 QLGKRG-----KKSQTLSELDDEAGFVSYLDKENMMWIGLSKEKQ-----STTIPYV 170  
 QY 89 FMKGLQMLMADAKKARRIKTNMMKNHFKHQLPYREMEHLRQFRODVTKCLFLGITSIP 148  
 DB 171 LGQAMQIFISDQPLDGLVYTPF-----EAPYQ--EHFERKWLNAEKTLEFGSVNLQ 220  
 QY 149 ----FA-----NYLVFLMLTFPR-----QLLR-----HFWT 172  
 DB 221 SPSLFAVYEPYRVIIQGNLQIYVEELNLEKVLLENTPIPTKOLFQFVMDVPHLE 280  
 QY 173 PQOQDFDI---YHAFKQSHPEIISYLEKVIPLISDAGLRRLDCTKIQRGTHPAI 229  
 DB 281 HUKADMIDIVKTHAI--QSEDELVSITKET--LISFG-QYRME-----322  
 QY 230 HDLALRECFSNHPLGMQOLAHVKALSRMLLTSLPPLLRHLKTHVYIHOIDKA 289  
 DB 323 -NVSVLEV-----IGRDYKELSLDKSKALFINPYVLGOLIKR--ETDSTFAELLNKO 374  
 QY 290 LAKLIGOL--TAQEVKSACYLRLGINS 314  
 DB 375 RIKAAQOLLSTSDSTENICVAVGYSN 401

RESULT 5  
 US-09-487-370-2  
 Sequence 2, Application US/09487370  
 Patent No. 6294652

GENERAL INFORMATION:  
 APPLICANT: Wallis, Nicola G.  
 APPLICANT: Ingraham, Karen A.  
 APPLICANT: Ge, Yigong  
 APPLICANT: Holmes, David J.  
 APPLICANT: Zalacain, Magdalena  
 APPLICANT: Throup, John  
 APPLICANT: Biswas, Sanjoy  
 TITLE OF INVENTION: Response Regulator  
 FILE REFERENCE: GM10091  
 CURRENT APPLICATION NUMBER: US/09/487,370  
 CURRENT FILING DATE: 2000-01-18  
 PRIOR APPLICATION NUMBER: 09/139,064  
 PRIOR FILING DATE: 1998-08-25  
 PRIOR APPLICATION NUMBER: 60/060,714  
 PRIOR FILING DATE: 1997-09-09  
 NUMBER OF SEQ ID NOS: 9  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 2  
 LENGTH: 428  
 TYPE: PRF  
 ORGANISM: Streptococcus pneumoniae  
 US-09-487-370-2

Query Match 4.9%; Score 94; DB 4; Length 428;  
 Best Local Similarity 20.8%; Pred. No. 0.047;  
 Matches 68; Conservative 56; Mismatches 115; Indels 88; Gaps 16;

QY 29 QLRSGGLAWGAPRSSKHLSPKADVKNLMSYVYTKAKINCKYHREFLGRHPREYIYTI 88  
 DB 122 QLGKRG-----KKSQTLSELDDEAGFVSYLDKENMMWIGLSKEKQ-----STTIPYV 170  
 QY 89 FMKGLQMLMADAKKARRIKTNMMKNHFKHQLPYREMEHLRQFRODVTKCLFLGITSIP 148  
 DB 171 LGQAMQIFISDQPLDGLVYTPF-----EAPYQ--EHFERKWLNAEKTLEFGSVNLQ 220  
 QY 149 ----FA-----NYLVFLMLTFPR-----QLLR-----HFWT 172  
 DB 221 SPSLFAVYEPYRVIIQGNLQIYVEELNLEKVLLENTPIPTKOLFQFVMDVPHLE 280  
 QY 173 PQOQDFDI---YHAFKQSHPEIISYLEKVIPLISDAGLRRLDCTKIQRGTHPAI 229  
 DB 281 HUKADMIDIVKTHAI--QSEDELVSITKET--LISFG-QYRME-----322  
 QY 230 HDLALRECFSNHPLGMQOLAHVKALSRMLLTSLPPLLRHLKTHVYIHOIDKA 289  
 DB 323 -NVSVLEV-----IGRDYKELSLDKSKALFINPYVLGOLIKR--ETDSTFAELLNKO 374  
 QY 290 LAKLIGOL--TAQEVKSACYLRLGINS 314  
 DB 375 RIKAAQOLLSTSDSTENICVAVGYSN 401

RESULT 6  
 US-09-487-370-4  
 Sequence 4, Application US/09487370  
 Patent No. 6294652

GENERAL INFORMATION:  
 APPLICANT: Wallis, Nicola G.  
 APPLICANT: Ingraham, Karen A.  
 APPLICANT: Ge, Yigong  
 APPLICANT: Holmes, David J.  
 APPLICANT: Zalacain, Magdalena  
 APPLICANT: Throup, John  
 APPLICANT: Biswas, Sanjoy  
 TITLE OF INVENTION: Response Regulator  
 FILE REFERENCE: GM10091  
 CURRENT APPLICATION NUMBER: US/09/487,370  
 CURRENT FILING DATE: 2000-01-18  
 PRIOR APPLICATION NUMBER: 09/139,064  
 PRIOR FILING DATE: 1998-08-25  
 PRIOR APPLICATION NUMBER: 60/060,714  
 PRIOR FILING DATE: 1997-09-09  
 NUMBER OF SEQ ID NOS: 9  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 4  
 LENGTH: 428  
 TYPE: PRF  
 ORGANISM: Streptococcus pneumoniae  
 US-09-487-370-4

Query Match 4.9%; Score 94; DB 4; Length 428;  
 Best Local Similarity 20.8%; Pred. No. 0.047;  
 Matches 68; Conservative 56; Mismatches 115; Indels 88; Gaps 16;

QY 29 QLRSGGLAWGAPRSSKHLSPKADVKNLMSYVYTKAKINCKYHREFLGRHPREYIYTI 88  
 DB 122 QLGKRG-----KKSQTLSELDDEAGFVSYLDKENMMWIGLSKEKQ-----STTIPYV 170  
 QY 89 FMKGLQMLMADAKKARRIKTNMMKNHFKHQLPYREMEHLRQFRODVTKCLFLGITSIP 148  
 DB 171 LGQAMQIFISDQPLDGLVYTPF-----EAPYQ--EHFERKWLNAEKTLEFGSVNLQ 220  
 QY 149 ----FA-----NYLVFLMLTFPR-----QLLR-----HFWT 172

Db 221 SESLFAYEPIRYVIOGNIQVIEENLLEKVLNLENTPRIPITKOLFIQVWDFHFE 280  
 QY 173 PROQTFELDI---VHAFKOSHPEITILEKVIPLISDAGLRRLDCTKIORGTHPA1 229  
 Db 281 HLKADMDIYKTHAI--OSFDELVSYIKET--LISFFG-QYRME----- 322  
 QY 230 HDIALRECCSNHPLGNMLOALHVKALSRAMELTSLPPLRLRHLKTHTVIHOLOKA 289  
 Db 323 -NVSVLEV-----IGRDYOKELSLKDISKALFINVYIGOLIKR--ETDSFAELMKQ 374  
 QY 290 LAKIGIQL--TAEVKSACYLGLNS 314  
 Db 375 RIKAQOULLSTSDSIENICTYAVGYSN 401

RESULT 7  
 US-08-451-715A-4  
 ; Sequence 4, Application US/08451715A  
 ; Patent No. 5801013  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tao, Jianshi  
 ; APPLICANT: Qul, Yan  
 ; APPLICANT: Houman, Fariba  
 ; APPLICANT: Shen, Xiaoyu  
 ; APPLICANT: Schimmel, Paul R.  
 ; TITLE OF INVENTION: Helicobacter Antinocyl-trNA Synthetase  
 ; NUMBER OF INVENTION: Proteins, Nucleic Acids and Strains Compising Same  
 ; CORRESPONDENCE ADDRESS: 67  
 ; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.  
 ; STREET: Two Millitia Drive  
 ; CITY: Lexington  
 ; STATE: Massachusetts  
 ; COUNTRY: USA  
 ; ZIP: 02173  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/451,715A  
 ; FILING DATE: 26-MAY-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Brook, David E.  
 ; REGISTRATION NUMBER: 22,592  
 ; REFERENCE/DOCKET NUMBER: CPI94-25  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-861-6240  
 ; TELEFAX: 617-861-9540  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 648 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-451-715A-4

Query Match 4.68; Score 87; DB 1; Length 648;  
 Best Local Similarity 18.18; Pred. No. 0.54;  
 Matches 50; Conservative 49; Mismatches 114; Indels 64; Gaps 8;  
 QY 28 LQGRSGIANGCAPRSSKHLSPKADVKNLMSYVYTKKALNGKTHRLGHPREFYLYT 87  
 Db 204 LSTRTSEWGIPLPKKN-DPKHIVYVWLDALINVASAG----- 243  
 QY 88 IFMGLQMLADAKKARIKTNMKRNHKKFQULPYREMEHLRQFQDYVKCLFLGIIISIP 147  
 Db 244 -YLGLDKNKAMHFCARHI--VGKILRFHAIYWP-----AFKSLNLP 284  
 QY 148 PFANVIVLLMYLPFRQLLRHFWT-----PKQDTFELDIYHAFKOSHPEITILEKV 201

Db 285 LF-----KOLCVHGWTFIEGVKMSKSLGNVDAQKLAIEYIEERFFLRE 331  
 QY 202 IPLISDAGLRWR--LTDLCIKIORGTHPAHDLALRECCSNHPLGNMLOALHVKALSR 259  
 Db 332 VPEGQDDDFPKKALIEFIENANLNNDGNLRLGNKRYFVYSLSAKITAAYYSKELEK 391  
 QY 260 AMLL-----TSLPPLRLRHLKTHTVIHOLOKALAK 292  
 Db 392 AHQILNANSEVPEKQMLHAELEFNVYDFLKLKAK 428

RESULT 8  
 US-08-750-152A-2  
 ; Sequence 2, Application US/08750152A  
 ; Patent No. 5977331  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ASAKURA, YOKO  
 ; APPLICANT: KIMURA, EIICHIRO  
 ; APPLICANT: ABE, CHIZU  
 ; APPLICANT: KAWAHARA, YOSHIO  
 ; APPLICANT: NAKAMATSU, TOSHIO  
 ; TITLE OF INVENTION: ALPHA-KETOGLUTARATE DEHYDROGENASE GENE  
 ; NUMBER OF SEQUENCES: 14  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.  
 ; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR  
 ; CITY: ARLINGTON  
 ; STATE: VA  
 ; COUNTRY: USA  
 ; ZIP: 22152  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: FastSeq Version 1.5  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/750,152A  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: NORMAN F. OBLON  
 ; REGISTRATION NUMBER: 24,618  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 703-413-3000  
 ; TELEFAX: 703-413-2220  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1257 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-750-152A-2

Query Match 4.48; Score 84; DB 2; Length 1257;  
 Best Local Similarity 29.58; Pred. No. 3.2;  
 Matches 38; Conservative 20; Mismatches 45; Indels 26; Gaps 8;  
 QY 258 SRAMLT-----SYLP--PLLNR--LKTHTTVIHOLOKALAKIGIG--QLTAEVK 304  
 Db 417 SRGLIDTNDPLNSVOPGMPYPRDRDIDITHSLTLMDDRTFSVGGFGKETMLIREVL 476  
 QY 305 S---ACYLRGLNS--THIGEDRCRTWLGEMLOISC--SLKAEILSLHNV-----VLL 351  
 Db 477 SRLAAVTLVAGSEYTHILDRERTWLDRLKMGMPYQAROKYILQKLNAEAFAENFL 536  
 QY 352 STNYLGRTR 360  
 Db 537 QTKYVGQR 545

```

APPLICANT: Panayiotatos, Nikos
TITLE OF INVENTION: Modified Ciliary Neurotrophic Factors
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Regeneron Pharmaceuticals, Inc.
STREET: 777 Old Saw Mill River Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/959,284
FILING DATE: 19921009
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kempler Ph.D., Gall M
REGISTRATION NUMBER: 32,143
TELECOMMUNICATION INFORMATION:
TELEPHONE: 914-347-7000
TELEFAX: 914-347-2113
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 200 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-07-959-284-8

Query Match          4.2%; Score 80; DB 1; Length 200;
Best Local Similarity 19.7%; Pred. No. 0.56;
Matches 34; Conservative 34; Mismatches 61; Indels 44; Gaps 8

QY 97 WADAKKARIKIKNMKHNKIKFQHPYREME---HROFRQDVTKCLFGIISIPFANY 152
      1: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 64 WSLTEAEHLQENLQAVR-TFHYLARLLIEDOOVHTPTBEGDQHIHLLLOVAFA-- 120
      1: : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 153 LVELLWYLRPPROHLIHNFTPKOOTDFLDIYNAFRKQSHPEIISYLEKYIPLISAGL-- 210
      1: : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 121 -----YQBELMILLEYKIPREADGMPIN-----VGDGGLPE 153
      1: : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 211 --RWRLTDLCTKIQRGTHPAIDITLIRECFESHHPICGMNOLALHYKALSRAM 261
      1: : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 154 KRLMGLKVL-QELSQWTVRSIHDLRYI---SSHQMSIALES-HYGAKDKOM 200
      1: : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
US-08-308-736A-8
; Sequence 8, Application US/08308736A
; Patent No. 5846935
; GENERAL INFORMATION:
; APPLICANT: Panayiotatos, Nikos
; TITLE OF INVENTION: MODIFIED CILIARY NEUROTROPHIC
; FACTORS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

```

INFORMATION FOR SEQ ID NO: 8:

Match	Conservative	Mismatches	Indels	Gaps
07	34	61	44	8

```
DB 64 WSELTAERLOEQLAYR-TFHVLLARLEDOQVHPTTEGDFHQAHTLLQVAF-- 120
QY 153 LVPLMLYEPROLLIHFHTPRQOQDFLDIYAFKROSPHETIISLVLEKVIPLISAGL-- 210
DB 121 -----YQIEEMILILEYKIPRNEADGMPIN-----VGGGLEE 153
QY 211 --RWRLDCTKIORGTHPAIDIALRECFSNHPLGMNOALHVKALSRAM 261
DB 154 KILMGLKVL-QELSWTVNSIHDRLV-----SSHOMGISALES-HYGAOKDM 200

RESULT 14
US-09-052-778-15
Sequence 15, Application US/09052778A
Patent No. 6060590
GENERAL INFORMATION:
APPLICANT: Bryant, Peter J.
APPLICANT: Kawamura, Kazuo
TITLE OF INVENTION: CHITINASE RELATED PROTEINS AND METHODS
TITLE OF INVENTION: OF USE
FILE REFERENCE: 07306/015001
CURRENT APPLICATION NUMBER: US/09/052,778A
CURRENT FILING DATE: 1998-03-31
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15
LENGTH: 452
TYPE: PRP
ORGANISM: Drosophila melanogaster
US-09-052-778-15
```

```
Query Match 4.1%; Score 79; DB 3; Length 452;
Best Local Similarity 22.7%; Pred. No. 2.5;
Matches 51; Conservative 27; Mismatches 75; Indels 72; Gaps 13;
```

```
QY 52 DVNLSYVYTK-----TKAINGKYHRLGHPREFYLTIF-----MKGLQMLM 97
DB 105 NVAILLSVGDKRIELDKAKELPNKYLELESPTGRTFVNTVSLVKTGSDGLDVM 164
QY 98 ADAR-KARRIKT---MMKHNINIKFHQ---LPRYREMEHLRQ-----FRQD----- 134
DB 165 QPKKPKKVKHSGIGMLMKGFVFGSDSIYDEKSEHKEQFALLRDVNAFRPNLL 224
QY 135 -----YKCLPLGIIISIPRANVYVPLMLYFPRQLLRHNTWTQKQTDPLD-----LY 183
DB 225 STTVLPNVNSLF---YDIPAVVNYLDFVNLGTF-----DEFTQRPDEIADYAPLY 274
QY 184 HARRKQSHPEI-----ISY-----LEKVIPLISDAGLRWRLTD 216
DB 275 EL--SERNPENVAOYKYWLRRNCPSKINVGATYGRWKLTLD 317
```

```
RESULT 15
US-08-911-853-31
Sequence 31, Application US/08911853
Patent No. 6048710
GENERAL INFORMATION:
APPLICANT: Gerltse, Gijbert
APPLICANT: Quax, Wilhelmus J.
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
TITLE OF INVENTION: EXPRESSION LEVELS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1013
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
```

```
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/911,853
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/699,092
FILING DATE: 16-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Glaister, Debra J
REGISTRATION NUMBER: 33,888
REFERENCE/DOCKET NUMBER: GC361-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-846-7620
TELEFAX: 650-845-6504
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 877 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-911-853-31
```

```
Query Match 4.1%; Score 79; DB 3; Length 877;
Best Local Similarity 20.3%; Pred. No. 6.7;
Matches 86; Conservative 63; Mismatches 152; Indels 122; Gaps 22;
```

```
QY 8 WARSAVWGSAY-----TPGHFVTR-----RLQGRSGLANGAPRSSKLHSPK--- 50
DB 49 WQASRPFSGVHYPLQARDNEPVRFRHLAESIRAOVEDFDLSFNPFAAMHQAPEVLG 108
QY 51 ---ADVKN-----LMGYVYTKTKAINGKXHRPL-GHNPREFYLY 86
DB 109 EYLADLRKIESRLYLVDFOCGQPIILDLVLSAMLERLAGNTRVILSGNHFQF--- 164
QY 87 TIFMGLQMLMDAKKARRIKTNMKNHNIKFHQLPYREMEHLRQFRODVTCLPLGIISI 146
DB 165 -----SLSKLQDNK-----LLCIDQHDMLRSLPV---QIQHLMNY-----LGG 199
QY 147 PEPANVYVFLMYLP-----RQLLRHNTPRKQOQDFLDIYAFKROSPHETIISY--- 197
DB 200 PELSPAAYVGSILMAKTEGMWGVKALMAH---ARGTEALQRFQF---GHPEIYDFG 252
QY 198 --LEKVIPLISDAGLRWRL-----TDLCTKI-QRGTHPAIDIALRECF---SNHPLG 245
DB 253 VVKKLSPLQHLDFLCSAIFERFDELCRDVDRSGSALLLEDAARLFMLPYDEYF-G 311
QY 246 MNQIALHVKALSRAMLTSYLPPLRLRRLKTHHTTVIHOIDKALAKLGIGQLTAQEVKS 305
DB 312 CYRYHA-----LLHDFLARLAVHKPQEVNAQLHRRBALALQGRDELALQHNOR 361
QY 306 ACYLGUNSTHGEDRCRWLG-----EWLQISCSLKEAELSLLHNVYLSNTVGL 357
DB 362 SGD-RALEQSMIGE-ACDQWRSGHFAEVLKWLFE---PLSEAELOXQSRLLVLMYLTALTL 416
QY 358 TRR 360
DB 417 SRR 419
```

```
Search completed: December 20, 2001, 10:11:05
Job time: 118 sec
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: December 20, 2001, 10:09:32 (without alignments)  
1070.939 Million cell updates/sec

Title: US-09-868-474-2

Perfect score: 1912  
Sequence: 1 MALSRCWARSAYWGSATP.....LSLLHNVLSTNYLGTRR 360

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

1:	/SIDS8_gcgcata/genseq/genseqp/AA1980.DAT *
2:	/SIDS8_gcgcata/genseq/genseqp/AA1981.DAT *
3:	/SIDS8_gcgcata/genseq/genseqp/AA1982.DAT *
4:	/SIDS8_gcgcata/genseq/genseqp/AA1983.DAT *
5:	/SIDS8_gcgcata/genseq/genseqp/AA1984.DAT *
6:	/SIDS8_gcgcata/genseq/genseqp/AA1985.DAT *
7:	/SIDS8_gcgcata/genseq/genseqp/AA1986.DAT *
8:	/SIDS8_gcgcata/genseq/genseqp/AA1987.DAT *
9:	/SIDS8_gcgcata/genseq/genseqp/AA1988.DAT *
10:	/SIDS8_gcgcata/genseq/genseqp/AA1990.DAT *
11:	/SIDS8_gcgcata/genseq/genseqp/AA1991.DAT *
12:	/SIDS8_gcgcata/genseq/genseqp/AA1992.DAT *
13:	/SIDS8_gcgcata/genseq/genseqp/AA1993.DAT *
14:	/SIDS8_gcgcata/genseq/genseqp/AA1994.DAT *
15:	/SIDS8_gcgcata/genseq/genseqp/AA1995.DAT *
16:	/SIDS8_gcgcata/genseq/genseqp/AA1996.DAT *
17:	/SIDS8_gcgcata/genseq/genseqp/AA1997.DAT *
18:	/SIDS8_gcgcata/genseq/genseqp/AA1998.DAT *
19:	/SIDS8_gcgcata/genseq/genseqp/AA1999.DAT *
20:	/SIDS8_gcgcata/genseq/genseqp/AA2000.DAT *
21:	/SIDS8_gcgcata/genseq/genseqp/AA2001.DAT *
22:	/SIDS8_gcgcata/genseq/genseqp/AA2001.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Query	Match	Length	DB	ID	Description
No.	Score					
1	1912	100.0	360	22	AAAB20492	Human cervical can
2	1071	56.0	203	22	AAAG62443	Human ribosomal pr
3	600	31.4	112	21	AAAF65193	Human 5' EST relat
4	337	17.6	162	21	AAAB58772	Lung cancer associ
5	124	6.5	586	16	AAW33891	Flea saliva protei
6	124	6.5	586	19	AAAM82365	Flea saliva protei
7	94	4.9	428	20	AAW95086	S. pneumoniae resp
8	94	4.9	428	20	AAW95087	S. pneumoniae resp
9	92	4.8	662	18	AAW20899	H. pylori cytoplasm
10	89.5	4.7	290	22	AAAG85022	Shrimp white spot
11	87	4.6	648	19	AAW75910	Helicobacter methi

12	86	4.5	2785	21	AAV57148	Human down-regulat
13	84.5	4.4	165	21	AA8A22761	Human ORF2031
14	84	4.4	462	22	AA89A615	Human protein sequ
15	84	4.4	580	20	AAV35191	Chlamydia pneumoni
16	84	4.4	1257	17	AA887628	Alpha-ketoglutaric
17	84	4.4	1257	19	AAW41781	B. lactofermentum
18	84	4.4	2150	21	AAV53898	Amino acid sequenc
19	84	4.4	5165	22	AA890617	Human secreted pro
20	83.5	4.4	500	16	AA84A561	Human prostacyclin
21	82.5	4.3	603	20	AAV35593	protein involved i
22	82.5	4.3	778	21	AA651229	Arbidopsis thailia
23	82.5	4.3	806	21	AA651228	Arbidopsis thailia
24	82.5	4.3	927	21	AA651227	Arbidopsis thailia
25	81.5	4.3	237	16	AA88A562	Human prostacyclin
26	81.5	4.3	709	17	AA899801	CRIL-7 nerve prote
27	80.5	4.2	742	21	AA8A2853	Human ORF2317
28	80.5	4.2	344	21	AA624909	Arbidopsis thailia
29	80.5	4.2	355	21	AA624908	Arbidopsis thailia
30	80.5	4.2	367	21	AA624907	Arbidopsis thailia
31	80.5	4.2	1545	19	AA636602	Rat sulphophytrea
32	80	4.2	200	15	AA853425	Human/rat chimerid
33	80	4.2	200	20	AA883340	Modified ciliary n
34	80	4.2	503	22	AAV72161	Human RNA metaboli
35	80	4.2	956	22	AA879523	Corynebacterium gl
36	80	4.2	1257	22	AA890996	C glutamylcum prote
37	79.5	4.2	388	21	AAV67237	Barracke fifth add
38	79.5	4.2	557	21	AA853400	Human colon cancer
39	79.5	4.2	873	21	AA805699	Human protein sequ
40	79	4.1	452	21	AAAB07182	Drosophila melanog
41	79	4.1	753	21	AAV66159	Arbidopsis thailia
42	79	4.1	877	19	AAV58328	Pseudomonas Orff s
43	79	4.1	877	21	AAV82601	Pseudomonas alcali
44	79	4.1	877	21	AA882235	Pseudomonas alcali
45	78.5	4.1	337	22	AA894947	Human protein sequ

## ALIGNMENTS

## RESULT 1

ID AAB20492 standard; Protein; 360 AA.

AC AAB20492

DT 21-JUN-2001 (first entry)

Human cervical cancer 1. protoncogene-encoded protein.

KW Cervical cancer 1 protooncogene; HCCR-1; oncogene; human; lung cancer; leukemias; lymphoma; kidney cancer; liver cancer

XX KW ovarian cancer; diagnosis; gene therapy.

Homo sapiens.  
XX

FT	key	Peptide
FT	key	Peptide

FT	Protein	26..360
14		7

FT	Domain	26.,.142
----	--------	----------

FT	Domain	143..162
----	--------	----------

FT	Domain	163..360
----	--------	----------

FT	Modified-site	313..315
----	---------------	----------

FT	Modified-site	42	Notes: "Oberbarrat
----	---------------	----	--------------------

ETT	Modified-site	48	/note="O-phosphoryl at
-----	---------------	----	------------------------

FT Modified-Site 34

100

FT	Modified-size	/note= "N-myristylated"
FT	38	
FT	/note= "N-myristylated"	
XX	WO200127149-A1.	
XX	19-APR-2001.	
XX	30-MAR-2000; 2000MO-KR00284.	
XX	15-OCT-1999; 59KR-0044811.	
XX	(KIMJ/) KIM J W.	
XX	Kim JW;	
XX	WPI: 2001-290710/30.	
XX	DR N-PSDB: AAF30809.	
XX	Novel human cervical cancer 1 protooncogene is useful in the diagnosis of various cancers, e.g., leukemia, lymphoma, kinder, liver, lung, ovary, and uterine cervix cancers -	
XX	Claim 3; Page 60-62; 67pp; English.	

CC The present sequence is that of the protein product of a novel  
 CC human cervical cancer I protooncogene, HCCR-1 (see AF038093).  
 CC The protooncogene was identified by differential display RT-PCR  
 CC analysis to isolate a clone that was expressed in cervical cancer,  
 CC metastatic tissue and CUMC-6 cervical cancer cells but not in  
 CC healthy tissue, and use of the isolated clone to screen a phase  
 CC lambda gtl1 human lung embryonic fibroblast cDNA library. The  
 CC HCCR-1 protein is markedly hydrophobic and has a characteristic  
 CC single membrane-spanning domain and pre-secretory signal peptide.  
 CC The presence of a C-terminal glycosylation site suggests that  
 CC HCCR-1 is a type II membrane protein. HCCR-1 protooncogene is  
 CC overexpressed in cervical cancer tissues, cervical cancer cell  
 CC lines and lung cancer cell lines. Overexpression is also observed  
 CC in leukemia, lymphoma, kidney, liver and ovarian cancers. The  
 CC protooncogene is used in claimed kits for the diagnosis of cancer.  
 CC Antisense sequences are used in a claimed process for treating or  
 CC preventing cancer in humans. Also claimed are a vector comprising  
 CC the protooncogene, a microorganism transformed with the vector,  
 CC especially *Escherichia coli* JM109/HCCR-1 (KTC0 06678P), and a  
 CC process for preparing HCCR-1 protein by culturing the host cells.

Query Match	100.0%;	Score 1912;	DB 22;	Length 360;
Best Local Similarity	100.0%;	Pred. No. 1.2e+200;		
Matches 360;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY	1	MAIRYOMASAVWGSVAVTPOHFAETRRLOJGRSGLAMGARRSKTLHSPRADVKNLMSY	60
Db	1	mairycvcaatsavagsavcpohvctrrlqdgysglawgapsskllhspkadvknllmsy	60
QY	61	VTKRAINGKYHRELGRHFRPRFYLLTYTFKKGLOMLADAKAKARIKTNNMKINIKENOL	120
Db	61	vtkckaingkyhrlqghfrfylltylftmkylqdmawadakarkrtknnmknikethql	120
QY	121	PYREMEHLROROBOVTCGLFGITSTIRPFANYVFLMYLFRPOLRIHNTWTKOQDFL	180
Db	121	pyremehlrtgrtgvctcslflgltstirpfanyvflmylfrpqlrlhntwcpqgdcfl	180
QY	181	DIYHAFKQSHPELISYLEKVLPDISDAGRNWRLTDLCTKIORGNPAINDIALRECFIS	240
Db	181	diyhaefkqshpeislylekvrpisdagrlwtldctkigqthpaindialrecofs	240
QY	241	NHPLGNMOLALNWKALSRMMLTSTYRPRLLNRRKLTHTTTHOJDLAKALIGIGLTA	3000
Db	241	nhplgmgldalnlwkalstramlstsyrrppllnrrktcthvlnqldakalagiglta	3000

[illegible]

RESULT	2
AA62443	
ID	AA62443 standard; Protein; 203 AA
XX	
AC	AA62443;
XX	
DT	05-SEP-2001 (first entry)
XX	
DE	Human ribosomal protein L14.22.

KW	Human: xkbosomal protein L14.22; malignant neoplasm; haemopathy;
KW	HIV infection; immunological disease; inflammatory disorder; cytostatic;
KW	haemostatic; virucide; immunomodulatory; antiinflammatory.
XX	
OS	Homo sapiens.

PN WO200138389-A1.

PD 31-MAY-2001.

PF 20-NOV-2000; 2000WO-CN00471.

PR 24-NOV-1999; 99CN-0124098.

PA (BIOR-) BIOROAD GENE DEV LTD SHANGHAI.

PI Mao Y, Xie Y;

DR WPI; 2001-355906/37.  
DR N-PSDB; AAH45403.

PT Human ribosomal protein L14.22 and encoded polynucleotide, used in diagnosis and treatment of malignant tumors, hemopathy, human immunodeficiency virus infection, immunological diseases and inflammation -

PS Claim 1; page 20-21; 30pp; Chinese

CC This invention relates to human ribosomal protein L14.22 and cDNA  
CC encoding it. The invention includes a vector containing the cDNA, a host  
CC cell transformed with the vector and an antibody directed against human  
CC ribosomal protein L14.22. The cDNA and protein sequences are used in the  
CC diagnosis and treatment of malignant neoplasms, haemophily, human  
CC immunodeficiency virus (HIV) infection, immunological diseases and  
CC various inflammatory disorders. Use of the protein and nucleotide  
CC sequence may result in cytostatic; haemostatic; virucide;  
CC immunomodulatory and antiinflammatory activity. This sequence represents  
CC human ribosomal protein L14.22.

Sequence 203 AA;

Query Match	56.0%;	Score 1071;	DB 22;	Length 203;
Best Local Similarity	100.0%;	Pred. No. 6.3e-109;		
Matches 203;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 158 MYRRQLLIHNFETPRQOQDFLDIYAFKQSHPELISYLEKVIPLISDAGLRMYLTDL 217  
1 MYLIPRGLIHTECPKQCFDIDLYHAFIKSHPELISYLEKVIPLISDAQIRWLTDI 60  
Db 218 CTKIQGSTHPIAHIIILRECFSESHYRSGMQLDALNHKALSRMLTSTYLRPLLNRK 277  
QY 61 CTKIQGSTHPIAHIIILRECFSESHYRSGMQLDALNHKALSRMLTSTYLRPLLNRK 277  
Db 61 CTKIQGSTHPIAHIIILRECFSESHYRSGMQLDALNHKALSRMLTSTYLRPLLNRK 277  
QY 278 THTTVIHQDLKALAKLGITGOLAEVNSACUYLRGLNSTHNGEGRCTYLGEMWQIGCSLK 337  
121 THTTVIHQDLKALAKLGITGOLAEVNSACUYLRGLNSTHNGEGRCTYLGEMWQIGCSLK 337

OY 338 EAEELSLHNVLLSTNLYGTRR 360  
 DB 181 eaelstllhmvllstnlygtrr 203

## RESULT 3

AAV65193  
 ID AAV65193 standard; Protein; 112 AA.

AC AAV65193;

DT 01-FEB-2000 (first entry)

DE Human 5' EST related polypeptide SEQ ID NO:1354.

XX Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;  
 KW gene therapy; chromosome mapping; upstream regulatory sequence;  
 KW forensic; location; development; protein synthesis; stability;  
 KW regulation; identification.

XX Homo sapiens.

PN WO9353051-A2.

XX 21-OCT-1999.

PF 09-APR-1999; 99WO-IB00712.

PR 09-APR-1998; 98US-0057719.

PR 28-APR-1998; 98US-0069047.

XX (GEST ) GENSET.

PA Dumas Malne Edwards J, Duclert A, Giordano J;

PI WPI: 2000-038446/03.

DR N-PSDB; AA42807.

XX Novel secreted protein 5' expressed sequence tag sequences used in  
 PT diagnostic, forensic, gene therapy, and chromosome mapping procedures

PS Claim 3; Page 761; 837pp; English.

XX AA42265 to AA43075 represent novel 5' expressed sequence tag (EST)

CC sequences, corresponding to human secreted proteins, corresponding to AA42265 to

CC AA43052. The 5' ESTs can be used for producing secreted human gene

CC products. They can be used to identify and isolate 5' untranslated

CC regions (UTRs) and upstream regulatory regions which control the

CC location, development stage, rate, and quantity of protein synthesis, as

CC well as stability of mRNA. The ESTs are also useful as probes for

CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can

CC also be used in forensic procedures to identify individuals, or in

CC diagnostic procedures to identify individuals having genetic diseases

CC resulting from abnormal gene expression. The products may also be used in

CC gene therapy protocols. The nucleic acids encoding signal peptides can be

CC used for directing extracellular secretion of a polypeptide or the

CC insertion of a polypeptide into a membrane, or importing a polypeptide

CC into a cell. The proteins encoded by the EST sequences may be useful in

CC treating a variety of human conditions. Secreted proteins have

CC therapeutic value, and the identification of new secreted proteins is

CC valuable. AA42249 to AA42264 and AAV64644 to AAV64650 represent

CC sequences used in the exemplification of the present invention.

XX Sequence 112 AA;

SO

Query Match 31.4%; Score 600; DB 21; Length 112;  
 Best Local Similarity 99.1%; Pred. No. 9.8e-58;  
 Matches 111; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 57 MSYVTKRAINGKHYRHLGRHFRFFLYTLTFKGLQMLADKKARITNMKKHNIK 116  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 1 msyvtckrakngkyhrflgrhfrfvylytlfmkgjqlmwadakkariktmmwkhnik 60

OY 117 FHOCPYREMEHLROFDVTKFLGIISIPPEFANYIVPLIMYFPPOLLIR 168  
 ||||||||||||||||||||||||||||||||||||||||||||||||

DB 61 fhqjpyremehlrqrdvkvkclflgllsppfanyivlflmylfppqjllir 112

## RESULT 4

AAB58272  
 ID AAB58272 standard; Protein; 162 AA.

AC AAB58272;

DT 14-MAR-2001 (first entry)

DE Lung cancer associated polypeptide sequence SEQ ID 610.

XX Human; lung cancer associated protein; neuroprotective; cytostatic;  
 KW cardioactive; immunomodulatory; muscular active; vulnerary;  
 KW gastrointestinal; nephrotropic; antiinfective; gynecological;  
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;  
 KW proliferative disorder; wound healing; infectious disease.

XX Homo sapiens.

OS WO20005180-A2.

XX 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05918.

PR 12-MAR-1999; 99US-0124270.

PR (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

PI Ruben SM;

DR WPI: 2000-587514/55.

DR N-PSDB; AAF18148.

XX Lung cancer associated gene sequences, referred to as lung cancer

PT antigens, useful for treatment, prevention, and diagnosis of disorders

PS such as lung cancer -

XX Claim 11; Page 1102; 1425pp; English.

XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer

CC associated proteins represented in AAB58106 - AAB58548. Lung cancer

CC antigens may have neuroprotective; cytostatic; cardioactive;

CC antagonists may have neuroprotective; cytostatic; cardioactive;

CC immunomodulatory; muscular active general; vulnerary; gastrointestinal

CC general; nephrotropic; antiinfective; gynecological; or antibacterial

CC activity. The invention also includes antibodies specific for the

CC protein or polynucleotide sequences. The lung cancer associated

CC polynucleotide sequences may be used for detection of lung cancer,

CC chromosome identification, as chromosome markers, and for numerous other

CC diagnostic or research purposes. The proteins may be used to treat

CC disorders such as neural, immune, muscular, reproductive,

CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative

CC disorders. The proteins may also be used in the treatment of wounds and

CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and

CC peptide AAB58549 are used in the course of the invention for the

CC identification and characterization of the polynucleotide and protein

CC sequences.

XX Sequence 162 AA;

SO

Query Match 17.6%; Score 337; DB 21; Length 162;  
 Best Local Similarity 90.3%; Pred. No. 1e-28;  
 Matches 65; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

QY 20 PGHVFTRRLQGRSLGAWGAPRRSKLHSPKADYKNLMSVVTKTAKINGVHRELGRHF 79
DB 27 pwtchpbelqgrsglwgagprsklnhsprkadvknmsvvtckvckvkngrhflgrhft 86
QY 80 PREYLYTTFMK 91
DB 87 prfxylytlfmk 98

```

## RESULT 5

```

AAW3891
ID AAW3891 standard; Protein: 586 AA.
XX
AC AAW3891;
XX
DT 14-APR-1998 (first entry)
XX
DE Flea saliva protein PfSPM(F)-586.
XX
KW Flea saliva protein; fspm(f); allergic dermatitis; therapy;
KM diagnosis; PfSPM(F)-586.
XX
OS Ctenocephalides felis.
XX
FH Key Location/Qualifiers
FT Misc-difference 379
FT /note="encoded by AWC"
XX
PN MO9737676-A1.
XX
PD 16-OCT-1997.
XX
PF 10-APR-1997; 97WO-US05959.
XX
PR 10-APR-1996; 96US-0630822.
XX
PA (HESK-) HESKA CORP.
XX
PI Hunter SW, Sim G, Weber ER;
XX
DR WPI: 1997-512409/47.
XX
DR N-PSDB: AAV92839.
XX
PT New flea saliva proteins - useful for treating allergic dermatitis
PT and as diagnostic reagents
XX
PS Example 3; Page 127-129; 179pp; English.
XX

```

This polypeptide comprises a non-full-length flea saliva protein, designated PfSPM(F)-586, encoded by a cDNA clone (see T92839), designated nifpm(f)-1758, isolated from a flea salivary gland library. PfSPM(F)-586 has a calculated mol. wt. of about 66,547 and a calculated pI of 4.80. A Genbank homology search revealed no significant homology between nifpm(f)-1758 and PfSPM(F)-586 and known nucleic acid or amino acid sequences. Novel nucleic acids (see T92830-27) and novel flea saliva proteins (W30480-91), their fragments or mimetopes can be used in claimed methods for treating allergic dermatitis in an animal, for testing if an animal is susceptible to, or has, allergic dermatitis and for desensitising a host animal to allergic dermatitis.

Sequence 586 AA;

Query Match 6.5%; Score 124; DB 18; Length 586;  
 Best Local Similarity 27.0%; Pred. No. 0.00014;  
 Matches 37; Conservative 29; Mismatches 69; Indels 2; Gaps 2;

```

QY 219 TKI-QRGTHPAIHDLALRCFSNHPGLGNOLQALHVKALSRAMLITSYLPPLLRHRLK 277
DB 34 tkmrsgsqasaneelmksklfede-tlcldslarpqqlvalcrvleislgtlnflrqlr 92
QY 278 THTTIVHOLDKALAKLIGQLTAQEVKSACYLRGINSHTIGEDRCRTWLGHWLQISCLK 337

```

```

DB 93 mklrslaaddknldqkegvsmlysevgqacaragrnmraygmpehrlrrqlcdwlnslnek 152
QY 338 EAELSLHNVVLSTN 354
DB 153 vppsllllsraimlpen 169

```

## RESULT 6

```

AAW82365
ID AAW82365 standard; Protein: 586 AA.
XX
AC AAW82365;
XX
DT 26-APR-1999 (first entry)
XX
DE Flea saliva protein PfSPM(F)-586.
XX
KW Flea saliva protein; PfSPM(F)-586; allergic dermatitis; allergen;
KM allergy; therapy; diagnosis; vaccine; ectoparasite.
XX
OS Ctenocephalides sp.
XX
FH Key Location/Qualifiers
FT Misc-difference 379
FT /note="encoded by AWC"
XX
PN W09845408-A2.
XX
PD 15-OCT-1998.
XX
PF 15-OCT-1997; 97WO-US18669.
XX
PR 10-APR-1997; 97WO-US05959.
XX
PA (HESK-) HESKA CORP.
XX
PI Weber ER;
XX
DR WPI: 1998-594480/50.
XX
DR N-PSDB: AAV73387.
XX
PT New isolated ectoparasite saliva genes - used to develop products
PT for the diagnosis, prevention, treatment and determining
PT susceptibility to allergic dermatitis
XX
PS Example 3; Page 99-101; 172pp; English.
XX

```

This non-full-length polypeptide comprises the 66 kDa flea saliva protein PfSPM(F)-586, as deduced from cDNA clone nifpm(f)-1758 (see AAV73387). It shows no significant homology to database sequences. The invention is directed to novel products and methods for isolating ectoparasite saliva proteins (ESPs). It provides ESps (see AAW82382-93), nucleic acid molecules encoding them, methods for their recombinant production, and therapeutic compositions for treating allergic dermatitis that comprise at least one ESP, as well as assay kits for testing if an animal has, or is susceptible to, allergic dermatitis, and a method of desensitising a host animal to allergic dermatitis using ESps. The ESps can also be used for the production of antibodies useful in diagnosis or in vaccines for flea saliva protein and the allergic dermatitis may be flea allergy dermatitis.

Sequence 586 AA;

Query Match 6.5%; Score 124; DB 19; Length 586;  
 Best Local Similarity 27.0%; Pred. No. 0.00014;  
 Matches 37; Conservative 29; Mismatches 69; Indels 2; Gaps 2;

```

QY 219 TKI-QRGTHPAIHDLALRCFSNHPGLGNOLQALHVKALSRAMLITSYLPPLLRHRLK 277

```

Db 34 tkmrsgagaganeimkfsklfedeltdlsarpqvalcorleisltgtnflrflr 92  
 QY 278 THTTTHLDKALAKIGTOLTAOEYKSCYRGLNTHIGEDRCRTWLGWLOISCLK 337  
 Db 93 mklrtaaddkmqkgvismcysevgacraratmympehrlrrqdedwinlsinek 152  
 QY 338 EAEISLLHNHVVLLSTN 354  
 Db 153 vppslllstralmpe 169

RESULT 7  
 AAM95086 standard; Protein; 428 AA.  
 AAM95086:  
 20-MAY-1999 (first entry)  
 S. pneumoniae response regulator polypeptide.  
 Response regulator; two component signal transduction system; TCSTS;  
 antibacterial; genetic immunisation; gene therapy; bacterial adhesion;  
 wound; body implant; bacterial infection; otitis media; conjunctivitis;  
 pneumonia; bacteremia; sinusitis; pleural empyema; endocarditis;  
 meningitis.  
 Streptococcus pneumoniae.  
 EP900846-A2.  
 10-MAR-1999.  
 02-SEP-1998; 98EP-0307054.  
 09-SEP-1997; 97US-0060714.  
 (SMIK ) SMITHKLINE BEECHAM CORP.  
 Biswas S, Ge Y, Holmes D, Ingraham K, Throup J;  
 Wallis N, Zalacain M;  
 WPI: 1999-155939/14.  
 N-PSDB: AAX26104.

New Streptococcus pneumoniae Response Regulator (RR) polypeptide and  
 polynucleotide - useful as diagnostic reagents and for prevention  
 and treatment of Streptococcus infections which cause  
 conjunctivitis, sinusitis and meningitis

Claim 4: Page 5; 56pp: English.

This represents a Streptococcus pneumoniae response regulator protein,  
 which is part of the two component signal transduction system (TCSTS),  
 host cells containing a vector comprising the nucleic acid are used for  
 the recombinant expression of the protein. The response regulator  
 sequences are useful for diagnosing a disease or susceptibility to a  
 disease related to the polypeptide levels. They can diagnose the stage  
 and type of infection. They are also useful for screening for compounds  
 which stimulate or inhibit polypeptide function. Agonists and antagonists  
 are useful for treatment of conditions associated with response regulator  
 imbalance, and are therefore potential antibacterial compounds. The  
 polynucleotides are useful for genetic immunisation and in gene therapy  
 and antisense CC sequences are useful for inhibition of expression of  
 the DNA. The response regulator polypeptides can prevent adhesion of  
 bacteria to matrix proteins, and are useful for use on wounds and body  
 implants to prevent bacterial infection. Anti-response regulator  
 antibodies induced by the polypeptide are useful for preventing or  
 treating infections, especially bacterial infections, and also for  
 isolating clones expressing the polypeptide. Diseases prevented,  
 diagnosed and treated include those caused by bacterial infection,  
 especially Streptococcus pneumoniae infections, which cause otitis  
 conjunctivitis, pneumonia, bacteremia, sinusitis, pleural empyema,

CC endocarditis and especially meningitis.  
 XX Sequence 428 AA:  
 SQ

Query Match 4.9%; Score 94; DB 20; Length 428;  
 Best local Similarity 20.8%; Pred. No. 0.17;  
 Matches 68; Conservative 56; Mismatches 115; Indels 88; Gaps 16;

QY 29 QLCRSGIAMGAPRSSKLIHSPKADYKNLMSSYVVTAKINCKYHNFELGRHPPREYLYTI 88  
 Db 122 qlgerg-----ksgqlsqeldeaqrivsyldkenwvlglskeqg-----stflpyyv 170  
 QY 89 FMKGLOLWADAKKARRIKTNMWNKINIKFHOPLYREMEHLRQFODVTKCLFGLTISPP 148  
 Db 171 lqdaqwffisdqpldglvrvpf-----eayq--ehferwklnaektlfygsvnaq 220  
 QY 149 ---FA-----NTLVPLMLFPR-----QLLR-----HPWT 172  
 Db 221 seslfayeplyrvlqgnlnqiveelnlekvlenrptipkqifqfmdvfnlfe 280  
 QY 173 PKQOTDFLDI---YHAFRKOSHPEIISLEYKVIPLISDAGLRWRLNLDCTKIQGTHPAI 229  
 Db 281 hlkadmtdivktlhal--gsfdelvyiket--liffg-qyrme----- 322  
 QY 230 HDITLRCFCFSPHPLGMNOLOALHVKALSRAMLITSTLPPLRLHKLKTHTYIHQLDKA 289  
 Db 323 -nvsvlev-----lgrdygkelslkalskalifnpyvjgqltkr--ecdstaalnkg 374  
 QY 290 LAKLIGQL--TAQEVKSACVLRGLNS 314  
 Db 375 rikaagqlllstdsienicyavgsn 401

RESULT 8  
 AAM95087 standard; Protein; 428 AA.  
 AAM95087:  
 20-MAY-1999 (first entry)  
 S. pneumoniae response regulator ORF sequence.  
 Response regulator; two component signal transduction system; TCSTS;  
 antibacterial; genetic immunisation; gene therapy; bacterial adhesion;  
 wound; body implant; bacterial infection; otitis media; conjunctivitis;  
 pneumonia; bacteremia; sinusitis; pleural empyema; endocarditis;  
 meningitis; ORF; open reading frame.  
 Streptococcus pneumoniae.  
 EP900846-A2.  
 10-MAR-1999.  
 02-SEP-1998; 98EP-0307054.  
 09-SEP-1997; 97US-0060714.  
 (SMIK ) SMITHKLINE BEECHAM CORP.  
 Biswas S, Ge Y, Holmes D, Ingraham K, Throup J;  
 Wallis N, Zalacain M;  
 WPI: 1999-155939/14.  
 N-PSDB: AAX26105.

New Streptococcus pneumoniae Response Regulator (RR) polypeptide and  
 polynucleotide - useful as diagnostic reagents and for prevention  
 and treatment of Streptococcus infections which cause  
 conjunctivitis, sinusitis and meningitis

PS Claim 19: Page6: 56pp; English.

XX This represents a Streptococcus pneumoniae response regulator protein,  
CC which is part of the two component signal transduction system (TCSTS).  
CC Host cells containing a vector comprising the nucleic acid are used for  
CC the recombinant expression of the protein. The response regulator  
CC sequences are useful for diagnosing a disease or susceptibility to a  
CC disease related to the polypeptide levels. They can diagnose the stage  
CC and type of infection. They are also useful for screening for compounds  
CC which stimulate or inhibit polypeptide function. Agonists and antagonists  
CC are useful for treatment of conditions associated with response regulator  
CC imbalance, and are therefore potential antibacterial compounds. The  
CC polynucleotides are useful for genetic immunisation and in gene therapy  
CC and antisense CC sequences are useful for inhibition of expression of  
CC the DNA. The response regulator polypeptides can prevent adhesion of  
CC bacteria to matrix proteins, and are useful for use on wounds and body  
CC implants to prevent bacterial infection. Anti-response regulator  
CC antibodies induced by the polypeptide are useful for preventing or  
CC treating infections, especially bacterial infections, and also for  
CC isolating clones expressing the polypeptide. Diseases prevented,  
CC diagnosed and treated include those caused by bacterial infection,  
CC especially Streptococcus pneumoniae infections, which cause otitis media,  
CC conjunctivitis, pneumonia, bacteremia, sinusitis, pleural empyema,  
CC endocarditis and especially meningitis.

XX Sequence 428 AA:

Query Match 4.9%; Score 94; DB 20; Length 428;  
Best Local Similarity 20.8%; Pred. No. 0.11;

Matches 68; Conservative 56; Mismatches 115; Indels 88; Gaps 16;

DB 29 QLRGSLAMGAPRSSKLLSPKADVKNLMSYVVTAKAINGKYRFLGRHPPRYLYTI 88  
122 qlgery-----ksgqlsgeldegfvsylgdkenwlglskekq-----stlpyyv 170  
QY FMKGLQMLMADAKKARRIKTMNMKNHIFHQLPYREMEHLROFRODYTKCLFLGISIP 148  
DB 171 lggawqfisdqldjlvtrf-----eapry--ehetrvklnckllfysvnlqg 220  
QY 149 -----FA-----NYLVLMYLFPR-----QLLR-----HFWT 172  
DB 221 sealfayeyryvllqgnlqveelnllekvlenltpirltkqlffqmdvfhle 280  
QY 173 PKQOTDFLDI---YHAFKOSHPELISYLEKVIRPLISDACLRLMDLCKTIOGTNHPAI 229  
DB 281 hlkadmdldvktihai--qsfdeivsyiket--liffg-qyrme----- 322  
QY 230 HDILALRECFSNHLLGMNOLOALHVKALSRAMLLTSTYRPLRLNRLKTHVTYIHQDKA 289  
DB 323 -nvsvslev-----lgrdygkelsldiskalfinpylqglktr--etdstfaellnkq 374  
QY 290 LAKLIGIQL--TAQEVKSACYLRGLNS 314  
DB 375 rikaaqqlilstdslenlcyavgyen 401

RESULT 9

AAW20999 AAW20999 standard; protein; 662 AA.

AC AAW20999;

DT 21-JUL-1997 (first entry)

DE H. pylori cytoplasmic protein, hp4el4522orf1.

XX Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;  
KM duodenal ulcer disease; chronic gastritis; diagnosis; envelope.  
XX Helicobacter pylori.

OS  
XX

PN WO9640893-A1.

XX 19-DEC-1996.

XX 06-JUN-1996; 96WO-US09122.

XX 01-APR-1996; 96US-0630405.

XX 07-JUN-1995; 95US-0487032.

XX (ASTR ) ASTRA AB.

PI Berglindh OT, Smith D, Mellgaard BJ.

DR WPI, 1997-052306/05.

DR N-PSDB; AAT68252.

PT Helicobacter pylori nucleic acid sequences and related

PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori

PS infection, and to detect Helicobacter

XX Claim 61: Page 1389-1390; 1481pp; English.

XX The present sequence is a H. pylori cytoplasmic protein involved in  
CC RNA translation and ribosome biogenesis.  
CC The protein may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds.  
CC The genomic sequence of H. pylori life cycle activators or inhibitors.  
CC DNA overlapping contigs generated by mechanically shearing the bacterial  
CC genome. The sequences were analysed for ORF of at least 160 nucleotides,  
CC and the predicted coding regions defined by computer evaluation. To  
CC identify likely H. pylori antigens for vaccine development, the amino  
CC acid sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
CC production, e.g. in E. coli hosts.

XX Sequence 662 AA:

Query Match 4.8%; Score 92; DB 18; Length 662;  
Best Local Similarity 17.7%; Pred. No. 0.53;  
Matches 49; Conservative 51; Mismatches 113; Indels 64; Gaps 8;

QY 28 IOLRSGIAMGAPRSSKLLSPKADVKNLMSYVVTAKAINGKYRFLGRHPPRYLYTI 87  
DB 210 lsirtsfewgdlpdkkm-dpxhvvyywldallnyasalq----- 249  
QY 88 IFMKGLQMLMADAKKARRIKTMNMKNHIFHQLPYREMEHLROFRODYTKCLFLGISIP 147  
DB 250 -ynglndknaheterahj--vgkdllrfhalyw-----afinsindp 290  
QY 148 PFANVLYVFLMYLFPRDLIRHFWT-----PKQOTDFLDIYHAFKOSHPELISYLEKV 201  
DB 291 lf-----kqlvhwgtlvgvkmkslgnvldaqklameygieelryfllire 337  
QY 202 IPLISAGLRMR--LMDLCTKIORTGTHPAIHDLALRECFSNHPLRGMMOALHVKALSR 259  
DB 338 vpfqgdgdfekkalverinanlndlglnlrlgmakkyfynkscklclaypkelek 397  
QY 260 AMLL-----TSYLRPLRLNRLKTHVTYIHQDKALAK 292  
DB 398 ahqldmansfvpkmqhlkaleelnfydflnklak 434

RESULT 10

AAAG85022 AAG85022 standard; Protein; 290 AA.

AC AAG85022;

DT 11-SEP-2001 (first entry)

XX Shrimp white spot Bacilliform virus (WSBV) protein 113.  
 DE Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;  
 XX antiviral agent; gene expression; antisense construct;  
 KW transgenic viral resistant shrimp.  
 KW White spot syndrome virus.  
 OS WO200138351-A2.  
 PN 31-MAY-2001.  
 PD 08-NOV-2000; 2000WO-US28888.  
 PF 24-NOV-1999; 99CN-0124717.  
 PR (PENY-) PE CORP NY.  
 PA (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANT C A.  
 XX (SINO-) SINGENOMAX CO LTD.  
 PI Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;  
 DR WPI: 2001-355877/37.  
 DR N-PSDB: AAH62802.  
 XX Primary nucleotide sequence of the shrimp white spot Bacilliform virus  
 PT (WSBV), useful for producing viral polypeptides that can be used to  
 PT screen for agents that are useful for treating WSBV infection -  
 PS Claim 1; Figure 3; 626pp; English.  
 XX The invention provides the primary nucleotide sequence of the WSBV genome  
 CC (AAH62889), predicted transcript sequences (AAH62689-AAH62839) and  
 CC encoded proteins (AAH62891-AAH62851) and oligonucleotide sequences  
 CC (AAH62840-63160) suitable for use as primers or probes. The nucleic acid  
 CC molecules and proteins of the invention are useful for diagnosis and  
 CC monitoring viral infection. In screens for antiviral agents and for  
 CC monitoring viral gene expression or activity during a treatment regimen.  
 CC The nucleic acid molecules are also useful as antisense constructs to  
 CC control viral gene expression in infected cells and tissues and to create  
 CC transgenic viral resistant shrimp.  
 CC Sequence 290 AA:  
 SQ  
 Query Match 4.7%; Score 89.5; DB 22; Length 290;  
 Best Local Similarity 26.9%; Pred. No. 0.3;  
 Matches 36; Conservative 21; Mismatches 42; Indels 35; Gaps 5;  
 QY 48 SPRADYKMLMSYVYTKAINKRYHFLGRHPRFYI-----LYTIFMK 91  
 Db 122 sqgkdvqevnklivkektlml-----frgfyspyikslpqlfgekektvstdlil 175  
 QY 92 GLQMLMADA-KKARIRIKTNMKNIKFHOUPYREMEHLROFRODYTKCLFLGISIPPA 150  
 Db 176 glfsadwmekilaiegmrksqlykwtvpraevehlkkkygd-----ivnplfk 227  
 QY 151 NY---LVPLMLTL 160  
 Db 228 synshcivldlyl 241  
 RESULT 11  
 AA75910  
 ID AAW75910 standard; Protein; 648 AA.  
 XX AAW75910;  
 AC AAW75910;  
 XX 12-NOV-1998 (first entry)  
 DT Helicobacter methionyl-tRNA synthetase.  
 XX

KW Helicobacter; aminocyl-tRNA synthetase; isoleucyl-tRNA synthetase;  
 KW methionyl-tRNA synthetase; leucyl-tRNA synthetase; valyl-tRNA synthetase;  
 KW lysyl-tRNA synthetase; seryl-tRNA synthetase; human gastric mucosa;  
 KW chronic gastritis; antimicrobial compound; antisense inhibition;  
 KW H. pylori; bacterium; enzyme.  
 OS Helicobacter pylori.  
 XX US5801013-A.  
 PN 01-SEP-1998.  
 PD 26-MAY-1995; 95US-0451715.  
 PF 26-MAY-1995; 95US-0451715.  
 PR 26-MAY-1995; 95US-0451715.  
 XX (CUBI-) CUBIST PHARM INC.  
 PA Houman F, Qiu Y, Schimmel PR, Shen X, Tao J;  
 DR WPI: 1998-494769/42.  
 DR N-PSDB: AAV53135.  
 XX DNA encoding Helicobacter amino-acyl-tRNA synthetase proteins  
 PT useful for recombinant production of the enzyme, anti-sense  
 PT constructs and hybridisation probes, and construction of tester  
 PT strains to test for inhibitors of the enzyme  
 PS Claim 7; Columns 61-66; 76pp; English.  
 XX This represents a Helicobacter methionyl-tRNA synthetase. The invention  
 CC provides nucleic acid sequences (AAV53134 to AAV53139) encoding  
 CC Helicobacter aminocyl-tRNA synthetases (hATTRNAs) (AAV5909 to AAV5914),  
 CC where the aminocyl-tRNA synthetase is isoleucyl-tRNA synthetase,  
 CC methionyl-tRNA synthetase, leucyl-tRNA synthetase respectively and at least  
 CC a portion of the hATTRNAs has catalytic activity or binding function. The  
 CC nucleic acids and recombinant cells containing the nucleic acids can  
 CC also be used for recombinant production probes in experiments to identify hATTRNAs  
 CC sequences. H. pylori is bacteria that infects human gastric mucosa,  
 CC leading to chronic gastritis. The amino acyl tRNA synthetases are used by  
 CC the bacterium during protein synthesis, and as such can be a target for  
 CC antimicrobial compounds. The tester strains can be used to screen for  
 CC such compounds, and antisense constructs based on the nucleic acids can  
 CC be used for antisense inhibition of the enzyme. The proteins themselves  
 CC can be used to raise antibodies, which in turn can be used for  
 CC purification and study of the enzyme. Screening of inhibitors of the  
 CC enzyme, as well as antisense antimicrobial activity through the nucleic  
 CC acids are novel approaches to inhibition activity against the bacteria,  
 CC as many present antibiotics used have side-effects and the bacteria are  
 CC becoming resistant to them.  
 CC Sequence 648 AA:  
 SQ  
 Query Match 4.6%; Score 87; DB 19; Length 648;  
 Best Local Similarity 18.1%; Pred. No. 1.8;  
 Matches 50; Conservative 49; Mismatches 114; Indels 64; Gaps 8;  
 QY 28 LQLRSGIAGAPRSSKLLSPRADYKMLMSYVYTKAINKRYHFLGRHPRFYI 7  
 Db 204 lsitrlsfewgiprkkm-dpkhvyvwdallnyasalg-----afmslnlp 243  
 QY 88 IFMKGLQMLMADAKKARIRIKTNMKNIKFHOUPYREMEHLROFRODYTKCLFLGISIP 147  
 Db 244 -ylngldnkmahfecaeh---vgkdllrhaalywp-----afmslnlp 284  
 QY 148 PFANYLVFLMYLFPRLIRHFWT-----PQQTDFLDIYHAFKQSHPELISYLEKV 201  
 Db 285 lf-----kqlcvhgwvwtlegvkmksklgnvldagklameygleelyfllire 331  
 QY 202 IPLISDAGLRMR--LTDLCTKIQGRTPAIHDLALRECSNHPLOMNOQLAHVKAISR 259

Db 332 vpfqgdgfskkaierianindgnllnrlgmakkyfmsylksakltaysskelek 391  
 QY 260 AMLL-----TSVLPPLRLRHLKTHHTYIHOIDKALAK 292  
 Db 392 ahqildanastvpmqjlkaleelfnyvdfinkltaik 428

## RESULT 12

AA57148  
 ID AAY57148 standard; Protein; 2785 AA.

AC AAY57148;

DT 28-FEB-2000 (first entry)

DE Human down-regulated in metastasis (DRIM) amino acid sequence.

KW Down-regulated in metastasis; DRIM; human; antimetastatic activity;  
 KM antibody; tumour; treatment; therapy.

OS Homo sapiens.

PN WO960116-A1.

PD 25-NOV-1999.

PF 17-MAY-1999; 99WO-EP03396.

PR 18-MAY-1998; 98EP-0303895.

PA (HOPE) ROCHE DIAGNOSTICS GMBH.

PI (ISIS-) ISIS INNOVATION LTD.

PI Weidle U, Tarin D;

DR WPI: 2000-053296/04.

DR N-PSDB; AA245136.

PT New polypeptide with antimetastatic activity, useful for therapeutic  
 PT compositions for tumour therapy

PS Claim 2; Page 42-49; 54pp; English.

CC This is the amino acid sequence of the human down-regulated in metastasis  
 CC (DRIM) protein. The protein has antimetastatic activity. The DRIM protein  
 CC is active in both its glycosylated and unglycosylated form, and can be  
 CC produced by recombinant technology in prokaryotic cells. DRIM mRNA is  
 CC strongly expressed in heart, skeletal muscle, pancreas, testis and ovary  
 CC tissues. The nucleotide and protein sequences can be used to create  
 CC anti-DRIM antibodies. The nucleic acids are useful in therapeutic  
 CC compositions, especially for treating tumours. They are also useful for  
 CC activating polynucleotides from the 5' untranslated region in gene  
 CC therapy.

XX Sequence 2785 AA;

Query Match 4.5%; Score 86; DB 21; Length 2785;  
 Best Local Similarity 19.8%; Pred. No. 19;  
 Matches 62; Conservative 49; Mismatches 94; Indels 108; Gaps 14;

QY 42 SSKHLSPKADVKMLSYVVTTKALNGKYHFLGRHFR-FYIIVTIPKGLQIMADA 100

Db 901 SSGKKKTRTAAKQIHLHVLVFSKFN-----pralyleeklyelqlllqgd 949

QY 101 KKAIRIKTN---MKKHNIKFOHPYRE-MEHL---RQFQDVTKCLFL---GIITPPRA 150

Db 950 qmqqktdlclmtykhp---hvyryrengrllledrsteelvhfsischnavvtahra 1006

QY 151 NYLVFLMYLF-----PROLLIRHFWT---PKQQTDFLD-----LYH 184

Db 1007 dlfpilmrlllygrmkntgsktgksgatmaivlrflagrqpeidqfildilfepvrh 1066

QY 185 AFRKSHPELIISTYLE-----KVLPISDAGLRKWRRLDCTKRGTHPAIHDLALRECF 239  
 Db 1067 fkngecsavvqavedldskvplgrqhg1----- 1097

QY 240 SNHPLGNMLOALVKAISRAMLTSYLPPLRLRHLKTHHTYIHOID----- 287  
 Db 1098 -----lnslsvlknshlsay1-pklqlllcmatavshldgreaklqlrfing 1147

QY 288 -KALAKIGTOLT 299

Db 1148 lknllrllgikmvt 1160

## RESULT 13

AA57148  
 ID AAB42267 standard; Protein; 165 AA.

AC AAB42267;

DT 08-FEB-2001 (first entry)

DE Human OREFX ORF2031 polypeptide sequence SEQ ID NO:4062.

KW Human; open reading frame; OREFX; detection; cytostatic; hepatotropic;  
 KW vulnary; antiproliferative; antiparkinsonian; neurotropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesteryl ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antineoplastic disease; coagulation;  
 KW thrombosis; contraceptive.

OS Homo sapiens.

PN WO200058473-A2.

PD 05-OCT-2000.

PF 31-MAR-2000; 2000WO-US08621.

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

PA (CURA-) CURAGEN CORP.

PI Shimkets RA, Leach M;

DR WPI: 2000-602362/57.

DR N-PSDB; AAC76476.

PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -

PS Claim 11; Page 3250-3251; 5507pp; English.

CC AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human OREFX open reading frames 1 to 3161. The OREFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antirheumatic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antineoplastic; antibacterial; antiviral; antineoplastic;



```

XX 03-JUN-1999.
PD
XX
XX 20-NOV-1998; 98WO-IB01890.
PF
XX 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97FR-0014673.
XX
XX (GEST ) GENSET.
PA
XX
XX Griffais R;
PI
XX WPI; 1999-357842/30.
DR
XX
XX Genome sequence of Chlamydia pneumoniae
PT
XX
XX Page 1049-1050; Disclosure: 1912pp; English.
PS
XX
XX AAY34584-Y35879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see AAY31990) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.
XX
XX Sequence 580 AA;
SQ

```

```

Query Match 4.4%; Score 84; DB 20; Length 580;
Best Local Similarity 23.7%; Pred. No. 3.3;
Matches 44; Conservative 23; Mismatches 59; Indels 60; Gaps 8;

QY 178 DFLDIYNAFKKOSH-----PEIISYLEK-----VIPLI---SDAGLR 211
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 261 dldlvekrgefyfnpflpeiledlekgllysvndakcvfheafsfmvgksdgyn 320

QY 212 WRLDLCTKIQRGTHPAIHDLALRECEFSNPTGMNQLQALHVAKALSRAMILTSYLPPL 271
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 321 yatdlaamyriedhadrkllivld-----lg---qslhtglleataaagylqp91 370

QY 272 LRH-----RLKTHP-----TVIHOLDKALAKLGIGOLTAOEVRKSACY 308
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 371 fshvgfglvdpggkklktsrgenvklrelltdaikaeealre-hrpeltdaigerap 429

QY 309 LRGLNS 314
   : | : | :
Db 430 vlgina 435

```

Search completed: December 20, 2001, 10:11:37  
Job time: 125 sec



Db 241 NHPGMMQOLAHVKALSRAMLTSYLPRLHRLKHTHTVTHQDLKALAKIGTQLTA 300  
 QY 301 OEVSACYLGLNSTHIGEDRCRTWLGEMLOISCSLKEAEELSLHNVLLSTNYLCTRR 360  
 Db 301 OEVSACYLGLNSTHIGEDRCRTWLGEMLOISCSLKEAEELSLHNVLLSTNYLCTRR 360

## RESULT 2

QY3X3  
 ID 09Y3X3 PRELIMINARY: PRT: 100 AA.

DT 01-NOV-1999 (TREMblrel, 12, Created)  
 DT 01-NOV-1999 (TREMblrel, 12, last sequence update)  
 DT 01-JUN-2001 (TREMblrel, 17, last annotation update)  
 DE HYPOTHETICAL 11.3 KDA PROTEIN (DKFZP586A011.PROTEIN).

GN DKFZP586A011.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=UTERUS;  
 RA Koehler K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.,  
 RL Submitted (Mar-1999) to the EMBL/GenBank/DBJ databases.

RN (2)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LUNG CARCINOMA;  
 RA Strausberg R.,  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AL050286: CAB43387.1;  
 DR EMBL: BC000395: AAH00395.1; -;  
 KW Hypothetical protein.

SQ SEQUENCE 100 AA: 11318 MW: 4C5179D6F75156C7 CRC64;

Query Match 27.2%; Score 520; DB 4; Length 100;  
 Best Local Similarity 100.0%; Pred. No. 6e-41;  
 Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 MLTSTYLPPLRLHRLKHTHTVTHQDLKALAKIGTQLTAOEVSACYLGLNSTHIGED 320  
 Db 1 MLTSTYLPPLRLHRLKHTHTVTHQDLKALAKIGTQLTAOEVSACYLGLNSTHIGED 60

QY 321 RCRTWLGEMLOISCSLKEAEELSLHNVLLSTNYLCTRR 360  
 Db 61 RCRTWLGEMLOISCSLKEAEELSLHNVLLSTNYLCTRR 100

RESULT 3  
 P91617 PRELIMINARY: PRT: 436 AA.

ID P91617  
 AC P91617

DT 01-MAY-1997 (TREMblrel, 03, Created)  
 DT 01-MAY-1997 (TREMblrel, 03, last sequence update)  
 DT 01-MAR-2001 (TREMblrel, 16, last annotation update)  
 DE ANON-66DB PROTEIN.

GN CG5989 OR ANON-66DB.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;  
 RN (1)  
 RP SEQUENCE FROM N.A.

RC STRAIN=SHAHINAV (WILD TYPE); TISSUE=WHOLE FLY;  
 RA MEDLINE=9737971; Pubmed=9236770;  
 RA Seebode-larsen S., Urbanczyk Mohebi B., Lambertson A.;  
 RT "The Drosophila ribosomal protein L14-encoding gene, identified by a  
 RT novel Minute mutation in a dense cluster of previously undescribed  
 RT genes in cytogenetic region 6D-";  
 Mol. Gen. Genet. 255:141-151(1997).

DR EMBL: Y10015: CAAT1122.1; -;  
 DR FlyBase: FBgn0017429; CG5989;  
 SQ SEQUENCE 436 AA: 50299 MW: 19A5EC0B5D71A5C CRC64;

Query Match 17.9%; Score 342; DB 5; Length 436;  
 Best Local Similarity 26.6%; Pred. No. 1.3e-23;  
 Matches 81; Conservative 67; Mismatches 148; Indels 8; Gaps 4;

QY 55 NLMSYVTKKAIKNGYHFLGNHPRFYLITTKMGLQIMADAKKARIK--TNMK 112  
 Db 125 NMDDYIFTRFNVKNDVLEKNEPKAMQLYRVEDVKQFFGDKRFLAIANDSP 184

QY 113 HNIKFMQLPRYEMENHROPDVTCLFLGITSIPFANVYLFMLYFPRLHRTHT 172  
 Db 185 OGIR--ALNKGLELYMQPRDMKVAPRALICSLPVMGYAFPLVETYPSPFLAHWT 242

QY 173 PKQGTPELDIYAFKQSPHELIISTLEKVIPLISDAGLRWRLDCTKIQGTAPAHDI 232  
 Db 243 PQORSEFQSYTKRKRLCNKKDVFRLQAKLAKATASHPKNSAFADILQIGSGTPTPEML 302

QY 233 LALRECFNHPGLMNOQLAHVKALSRAMLTSYLPRLHRLKHTHTVTHQDLKALAK 292  
 Db 303 IDVKDIFAGCPYSLGMSKHYRNLVNLHGLPST--FKNHLRHEHAFVTHWDQATIR 359

QY 293 L-CIGQLTAOEVSACYLGLNSTHIGEDRCRTWLGEMLOISCSLKEAEELSLHNVLL 351  
 Db 360 EGGVHNLTPDALRYSCYGLNSTHIGEDRCRTWLGEMLOISCSLKEAEELSLHNVLL 419

QY 352 STNY 355  
 Db 420 GYNH 423

RESULT 4  
 Q9VSM4 PRELIMINARY: PRT: 436 AA.

ID Q9VSM4  
 AC Q9VSM4

DT 01-MAY-2000 (TREMblrel, 13, Created)  
 DT 01-MAY-2000 (TREMblrel, 13, last sequence update)  
 DT 01-MAR-2001 (TREMblrel, 16, last annotation update)  
 DE ANON-66DB PROTEIN.

GN CG5989 OR ANON-66DB.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;  
 RN (1)  
 RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;  
 RA MEDLINE=20196006; Pubmed=10731132;  
 RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George K.Y., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Egholm A.E., Garay N.S., Gelfand W.M., Glasser K.,  
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Fleischmann W.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kravitz S., Kulp D., Lai Z.,



```

Db 135 -VPTFLVFPGLFLIEIPVFLIPRYVLPDCYWRDQSKYLTQLATSRAYGMILHLS 193
OY 200 -KVPLISDAGLRWRLTDCTKIORGTHPAIHDLALRECFSNHPLGMNOALHVKALS 258
Db 194 YHNDNLISDORAKORHALVQVSEGVPSYDLSLHLPLFRD-----S 237
OY 259 RAMLITSYLPPLLRHRLKHTTGV-----IHOLDKALAKG-IGOLT 299
Db 238 SSLRLISDLNPNVLRSLCNATFAIVPQTKSMALRSLTRMDLFLRLDRKLRDEDLHLKLS 297
OY 300 AOEKKSACVLRGNSHTIGEDRCRTWLGEMWLOISCSLKEAEELSLHLNHYLSTNLGTR 359
Db 298 PLHLDTATLMRGDLSLSPDANKRYFLQHWLQVTRCSADDTVMFLHAWLMSFNSETK 357

RESULT 7
095202 PRELIMINARY: PRT: 739 AA.
AC 095202;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE LEUCINE ZIPPER-EF-HAND CONTAINING TRANSMEMBRANE PROTEIN 1.
GN LETM1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Endele S., Pak S.-J., Fuhry M., Zabel B., Winterpacht A.;
RT "LETM1, a novel leucine zipper and EF-hand containing transmembrane
RT protein gene located on chromosome 4p16.3."
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: NO EF-HAND FAMILY.
EMBL: AF061025; AADI3138.1; -.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF00036; ehfand.1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
KW Calcium-binding; Transmembrane.
SQ SEQUENCE 739 AA; 83353 MW; 942E9138F29D94F CRC64;

Query Match 9.4%; Score 180.5; DB 4; Length 739;
Best Local Similarity 23.8%; Pred. No. 2.8e-08;
Matches 88; Conservative 60; Mismatches 160; Indels 61; Gaps 14;

OY 3 LSRVCAARSVMGSAATPGHFYTRQLQGR-----SGLAWGAPRSSKHL-----LS 48
Db 83 VSRAPWTSTVCGFAVAGPQCLPYRGHSSRPVDRDSVVERKSLSKDKKKLEEGGPVYS 142
OY 49 PKADV--KNLSYVYTKTKAINGKYHRELFGRHFFRYTILYTFMKGLQMLWADAKKAR 105
Db 143 PRAEVVVKKSLGQGVLDLKHV---YH-----GRLMLMDITKIAAR 180
OY 106 IKTNNMKHNKIFQOLPYREMEHLKQFRODYTCLEGLTISPPRANVLYFLMYLPPROL 165
Db 181 M--LWR-ILNGISLTRRRROFLRICADLFRVLPFLVEVVPFMFLPVAVKLP--NM 235
OY 166 LHFHWTPROQDFLDIYHAFKROSHP--EISYLEKVI---PLISDGLRRRLDLCIK 220
Db 236 LSTFTFQSLKEERL-----KKELRVKLELAKFLQDTLEEMALKKAKKAGSATKDFSVF 289
OY 221 IGR---GTHPAIHDLALRECFSNHPLGMNOALHVKALSRAMLITSYLPPLLRHL 276
Db 290 FOKIRETGRSPNEEIMRSKLFEDF-LTLDNLTROPVALCKLLELOISIGNNLRQL 348
OY 277 KHTTYIHOLDKALAKLGIGULTAOEVSACVLRGNSHTIGEDRCRTWLGEMWLOISCSL 336
Db 349 TMRLSIKADDLIAEEGVDSINVKELQAACRARGMALGVTEDRQLKQMDLHLH- 407
OY 337 KEAEELSL 345

```

```

Db 408 QETPSTLI 416

RESULT 8
092210 PRELIMINARY: PRT: 738 AA.
AC 092210;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE LEUCINE ZIPPER-EF-HAND CONTAINING TRANSMEMBRANE PROTEIN 1.
GN LETM1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Endele S., Pak S.-J., Fuhry M., Zabel B., Winterpacht A.;
RT Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF061026; AADI3139.1; -.
DR InterPro: IPR002048; EF-hand.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 738 AA; 82988 MW; 5074CE630940729 CRC64;

Query Match 8.9%; Score 170; DB 11; Length 738;
Best Local Similarity 24.4%; Pred. No. 2.6e-07;
Matches 65; Conservative 54; Mismatches 125; Indels 22; Gaps 9;

OY 89 FMKGLQMLWADAKKARRTKTNMKHNKIFQOLPYREMEHLKQFRODYTCLEGLTISPP 148
Db 163 YHNGRLLMIDTKIAAR--LWR-ILNGHTLTRRRROFLRICADLFRVLPFLVAVVP 218
OY 149 FANYLVFLMYLFPKQDLIRHWTPKQOTDFLDIYHAFKROSHP--EISYLEKVI--P 203
Db 219 FMEFLPVAVKLP--NMLPSTFETQSIEERL-----KKELRVKLELAKFLQDTLEMA 271
OY 204 LISDGLRWRLDLCIKR-----GTHPAIHDLALRECFSNHPLGMNOALHVKALSR 259
Db 272 LKNKAAGKMATKDSAFQKIRETGERSPNEIMRFSKLFEDF-LTLDNLTROPVALCK 330
OY 260 AMLTYSYLPPLLRHRLKHTTYIHOLDKALAKLGIGULTAOEVSACVLRGNSHTIGE 319
Db 331 LLELOISTNNFLKQOLMRSLIKADDKLISEGVDSLYKELQAACRARGMALGVTE 390
OY 320 DRCRTWLGEMWLOISCSKEAEELSL 345
Db 391 DRLKQOLKQMDLHLH-HEIPTSLI 415

RESULT 9
004471 PRELIMINARY: PRT: 398 AA.
AC 004471;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE SIMILAR TO SACCCHAROMYCES HYPOTHEICAL PROTEIN P9642.2.
GN F5114.7.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Rosidae;
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CV. COLUMBIA;
RA Vysotskaya V.S., Osborne B.I., Toriumi M., Yu G., Oji O., Shen Y.K.,
RA Buehler E., Conway A.B., Conway A.R., Dewar K., Feng J., Kim C.,
RA Kurtz D., Li Y., Sun H., Davis R.W., Ecker J.R.,
RA Federspiel N.A., Theologis A.;

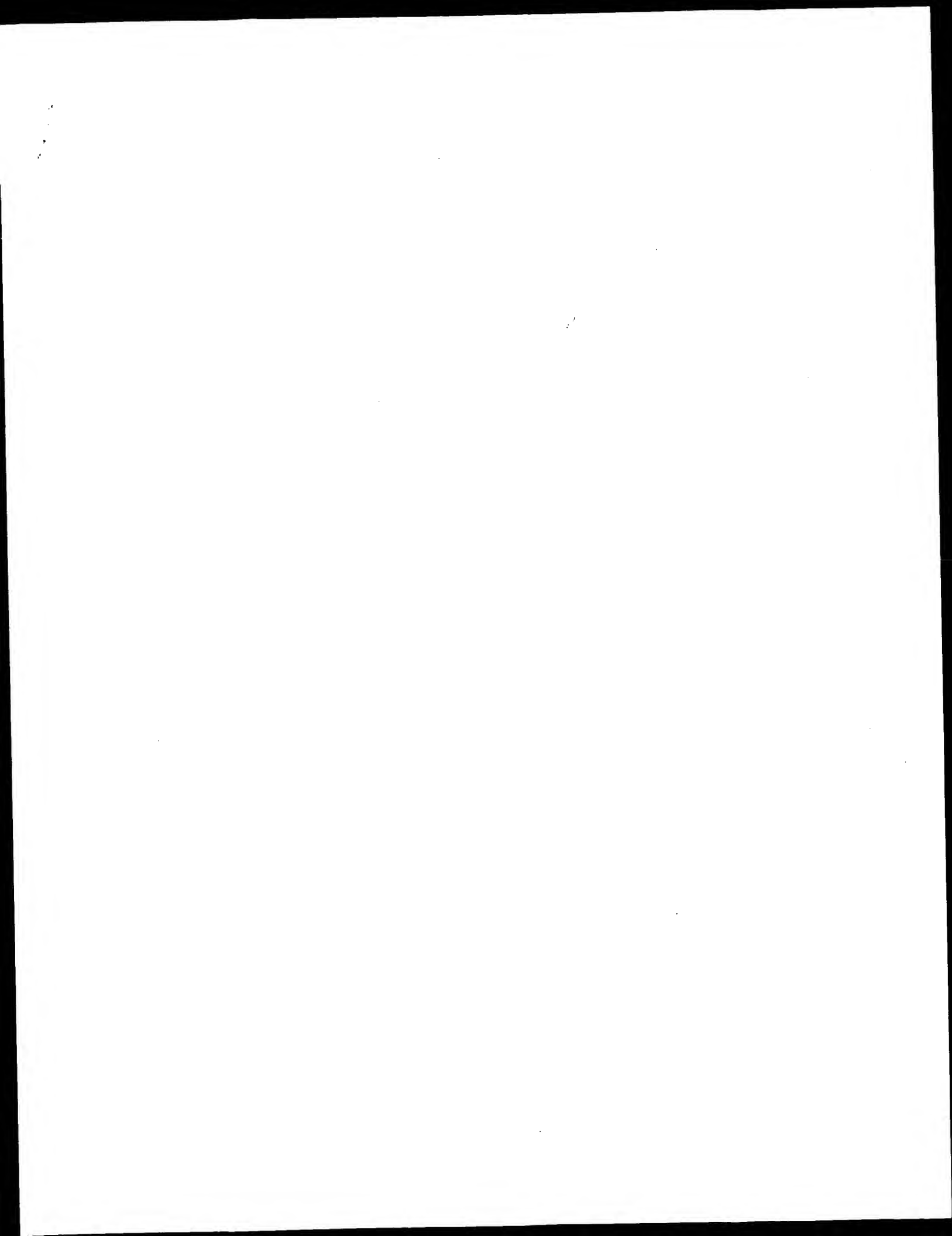
```

Oy 40 PRSSKHLSPK-ADVKNLMISVYTKTAKNGKYHFFLGRHFPRFYLLTYTFPKGLDMMA 98  
| | : : | : | : | :  
Db 90 PTPSKVAEPKOYAAEETKPTTVKKPST---WQRVKGG-----VLH--FNDGTIKLLGV 138

QY    227 PAIHDLIAIECFNHPHJCMNOLOALHYKALSRAMLLTSTYPPPLRLRKLHTTIVHQL 286B  
     :  
Db    240 VSNHEELIKSKLPEDE-TILDLSNGQLBSRCIRLMSTINSLSPELLRRQNLMKIRELKAD 298B

ID	Q9M122	PRELIMINARY:	PRT:	755 AA.
AC	Q9M122:			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	HYPOTHETICAL 85.7 KDA PROTEIN.			
GN	F24G16.90.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	Eurosids II; Brassicales; Brassicaceae; Arabidopsi.			
OX	NCBI_TaxID=3702;			
RP	SEQUENCE FROM N.A.			
RA	D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,			
RA	Lencke K., Mayer K.F.X., Queller F., Salanoubat M.;			
RL	Submitted (Feb-2000) to the EMBL/Genbank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	EU Arabidopsis sequencing project;			
RA	Submitted (Feb-2000) to the EMBL/Genbank/DBJ databases.			
CC	-I- SIMILARITY: TO PF-HAND FAMILY.			
DR	EMBL: AL138647; CAB/5801.1; .			
DR	InterPro: IPR02048; EF-hand.			
DR	Pfam: PF00036; ehand.			
DR	PROSITE: PS00018; EF_HAND; UNKNOWN_1.			
KW	Calcium-binding; Hypothetical protein.			
SQ	SEQUENCE 755 AA; 85734 MW; F631EF9307AF6108 CMC64.			
Query Match 7.1%; Score 136.5; DB 10; Length 755;				
Best Local Similarity 19.7%; Pred. No. 0.00036;				
Matches 50; Conservative 56; Mismatches 122; Indels 25; Gaps				
QY	92 GLOMMAQAKKARRIKTKMMKNKIKFHOQLPYREMHILQFRODYKICLFLGISIPPRAN 151			
DB	218 GIKLMAADTRISSRLKL---AGSKLSRRERQQLTRTTADIRLPAFAVITLPPWE 273			
QY	152 YLVFLIMLVFPROLIRHFHTFKQOTDFLDIYHARKQ--SHPEITSYLEKVIPLISDAG 209			
DB	274 FLPLPVELKLPNML-----PSTQDMKKEBEAKRLRLARIEAKFLQETARMAKEV 326			
QY	210 LRMV-----LTDCTKIQRSTHRAINDIALRRCFSNHPGNNOLALHWKALS 259			
DB	337 KHSRGEVKAQAEDEDFLDKRVRSQIVHNDELGLFAFLF-NDELTLLNISRPRLVSCK 385			
QY	260 AMLLTSLPPLRLRHRLHTHTTHVHQLDLAKALIGQLTQEVVSACYLGRNSTHIGE 319			
DB	386 YMGISPYGTDAVLRMLMKRLKRLKSTREDDKLIRAEVSLSLSEHLREDCREKMGGL-VSV 444			
QY	320 DRCRTWGEMLQIS 333			
DB	445 EEMRQQLRDMWDL 458			
RESULT 14				
008179	ID	PRELIMINARY:	PRT:	573 AA.
AC	008179:			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	CHROMOSOME XV READING FRAME ORF Y01027C.			
GN	Y01027C.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Habbib B., Hattenhorst U., Hollenberg C.P., Ramezani Rad M.;			
RA	Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.			





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2001, 16:09:03 ; Search time 1816.46 Seconds

(without alignments)  
12529.623 Million cell updates/sec

Title: US-09-868-474-1

Perfect score: 2118

Sequence: 1 ctgtgaagatgctctcc.....aaaaaaaaaaaaaaaaaaaaa 2118

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 segs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

EST:\*

1: em\_estfun:\*

2: em\_esthum:\*

3: em\_estlin:\*

4: em\_estom:\*

5: em\_estpl:\*

6: em\_estba:\*

7: em\_estro:\*

8: em\_estov:\*

9: em\_hic:\*

10: gp\_est1:\*

11: gp\_est2:\*

12: gp\_hic:\*

13: gp\_gss:\*

14: em\_gss\_fun:\*

15: em\_gss\_hum:\*

16: em\_gss\_inv:\*

17: em\_gss\_pln:\*

18: em\_gss\_pro:\*

19: em\_gss\_tod:\*

20: em\_gss\_vrt:\*

21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	913.6	43.1	951	10	AL542968
C 2	848.2	40.0	885	10	AL576116
C 3	784.4	37.0	864	10	AL576116
C 4	772.6	36.5	801	10	AL576116
5	757.4	35.8	863	11	AL576116
6	738.2	34.9	849	11	AL576116
7	736.6	34.8	819	11	AL576116
8	731.4	34.5	804	11	AL576116
9	717	33.9	804	11	AL576116
10	713.4	33.7	850	11	AL576116
11	711.8	33.6	844	11	AL576116
12	700.2	33.1	967	11	AL576116

13	689.4	32.5	702	11	BI255100
14	686	32.4	714	11	BI255100
15	681	32.2	702	10	BI255100
16	672.4	31.7	687	11	BI255100
17	664	31.4	740	11	BI255100
18	658.2	31.1	694	10	BI255100
19	655	30.9	876	10	BI255100
20	649.8	30.7	889	11	BI255100
21	642.8	30.3	1085	11	BI255100
22	635.2	30.0	672	10	BI255100
23	634	29.9	827	11	BI255100
24	617.8	29.2	618	11	BI255100
25	606.4	28.6	941	11	BI255100
26	604.4	28.5	615	10	BI255100
27	597	28.2	713	10	BI255100
28	591.4	27.9	601	10	BI255100
29	576	27.2	649	10	BI255100
30	571.4	27.0	790	11	BI255100
31	558	26.3	605	10	BI255100
32	550	26.0	665	10	BI255100
33	548.4	25.9	1111	10	BI255100
34	542.4	25.6	548	10	BI255100
35	539.8	25.5	748	11	BI255100
36	539.4	25.5	561	10	BI255100
37	534.8	25.3	591	11	BI255100
38	532.2	25.1	591	11	BI255100
39	530.6	25.1	791	10	BI255100
40	523.8	24.7	725	11	BI255100
41	522.6	24.7	731	11	BI255100
42	520.8	24.6	795	10	BI255100
43	518	24.5	529	10	BI255100
44	518	24.5	946	11	BI255100
45	517.4	24.4	530	10	BI255100

#### ALIGNMENTS

RESULT 1	AL542968	951 bp	EST	16-FEB-2001
LOCUS	AL542968	951 bp	EST	16-FEB-2001
DEFINITION	AL542968 LTI_FL002_P11 Homo sapiens cDNA clone CS0DE013YH04 3 prime			
ACCESSION	AL542968			
VERSION	AL542968.1	GI:12875446		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	Li, W.B., Gruber, C., Jesse, J., and Polayes, D.			
TITLE	Full-length cDNA libraries and normalization			
JOURNAL	Unpublished (2001)			
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.			

#### FEATURES

Source

1. 951

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="CS0DE013YH04"

/clone\_11p="LTI\_FL002\_P11"

/lab="host-DH10B"

/note="Organ: Placenta; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville , Maryland 20850, USA Fax : (1) 301 610 8371 Email :



```

Db      525 ACATCTGTGGGTTTTTTGTTGGTCTGTAGAAAAATTTGGTGCGTGTAAGAAACAGACACTCC 466
QY      1537 ttggcgtgagcactgtgtcgcagtgactggtgttcccccacactccttgcgt 1596
        |||||||
Db      465 TTGGGCTGGAGCACTGTGTGCMGTGCATGACTGGGTGTTCCCTGCATCCTTTCTGAT 406
QY      1597 atgacacaaatcaagttgttcttcttcttctcactcactgacatggcctaacact 1656
        |||||||
Db      405 ATGACCAAAATACAGATGTTGTTTGTGTCACCTTMACCTGGCAATGGGCTTAACACACTT 346
QY      1657 ctttccaaaccttgacacacttttctgagtgaacttgacaggaatattctatctg 1716
        |||||||
Db      345 CTTTTCAAACCTCTGAAACCTTTTCTGATGGTGAATCTGCAGAAATATCTATGG 286
QY      1717 aaaaagtaagagaagtaagaatgctcttgaccctctctaatttcttgccttacc 1776
        |||||||
Db      285 AAAAGATACAGAGAGTACAGATGCTCTTGACCCCTCCCTCAATGTTCTACCTTCAC 226
QY      1777 tctcattcttcttctggtcgtatatacagccctctgtgagatcctcactgtgcct 1836
        |||||||
Db      225 TCTGCATGTGCTTTCTGGGCTGTATTACAGCCVTCTGTGATCTTCAACTCTGCTGCCCT 166
QY      1837 ccaactgtatgcagcagtcacactgtaactgacagtgtgcctctctcgtggccatgat 1896
        |||||||
Db      165 CCACGTGTATGCAGCAGRCACACTGTAAGTACAGTSGCTGCTCTGTGGGCATGAT 106
QY      1897 caacactgtaagtaactaactgcccagcctggggaagataaggaagttcgtcaat 1956
        |||||||
Db      105 CACACCTGTAGTACTAATTACTGCCATTCCTGGGAGAAATCAGAGACTTCTGCATAGT 46
QY      1957 tagtaagttggttagctcttctgtgtgcatcagtaactagag 2001
        |||||||
Db      45 TACTAGTTGGGTTTGTAGCTTTGTGTGTCATCAGTACTTASAG 1

```

**RESULT 3**  
**AL516271/c**  
**LOCUS** AL516271 864 bp mRNA EST 13-FEB-2001  
**DEFINITION** AL516271 LTI\_NFL011\_NBC1 Homo sapiens cDNA clone CS0DA005YD18 3  
**ACCESSION** AL516271  
**VERSION** AL516271.1 GI:12779764  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
**REFERENCE** 1 (bases 1 to 864)  
**AUTHORS** Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
**TITLE** Full-length cDNA libraries and normalization  
**JOURNAL** Unpublished (2001)  
**COMMENT** Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
**FEATURES**  
 source  
 1..864  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CS0DA005YD18"  
 /clone\_11b="LTI\_NFL011\_NBC1"  
 /sex="male"  
 /tissue\_type="neuroblastoma cells"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
 was primed with a NotI-oligo(dT) primer. Five prime end  
 enriched, double-stranded cDNA was digested with Not I and  
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6  
 vector. Library is not normalized, but is the control for  
 the normalized libraries. Library was constructed by Life  
 Technologies. Contact : Feng Liang Life Technologies, a  
 division of invitrogen 9800 Medical Center Drive Rockville  
 , Maryland 20850, USA Fax : (1) 301 610 8371 Email :

```

f1iang@life.techn.com URL :
http://fulllength.invitrogen.com"
BASE COUNT      265 a      194 c      204 g      185 t      16 others
ORIGIN
Query Match      37.0%; Score 784.4; DB 10; Length 864;
Best local similarity 93.8%; Pred. No. 8.5e-126;
Matches 811; Conservative 13; Mismatches 40; Indels 1; Gaps 1;
QY      1134 atagcagtgacagagaacacacagcacttgccagcaagtcgtgtgtaactgtgtaagtggt 1193
        |||||||
Db      864 ATGCAGAGCGCAGAGAACAAACAGCACTTCCAGCAAAAGCTGTGTAACTGTAAGTGTGT 805
QY      1194 gggagagcagagagagagagcagggccatgggcttcaagcatgagacacacttggagact 1253
        |||||||
Db      804 GGRAGGCAAAAGAGAGAGACAGAGCCCATGGGCTTCACAGCTTAGMACACTGTGGGAACT 745
QY      1254 gaagacattccctcagcagctagaactgaacaaacccctctgtcgaaggtgtccgtgt 1313
        |||||||
Db      744 GCAAACTTCCCTTACAGCTAGAACTGAACAAACCTCTTGTGTAAGGGTGTCCGTRT 685
QY      1314 gagggtcatccctgtcccccataataactaataagtcgtgagacgtgagcagcctact 1373
        |||||||
Db      684 GAGGTGTCATCCTGTGCCCTTATATTAATGCTGGAACRGRGACAGCTCAACT 625
QY      1374 gggcttcaactgtgagtgttcgaatgaatgtccttaggaagtcagcttggcccaagtg 1433
        |||||||
Db      624 GGGCTTTAACTGTGATGTTGTCAGTTCATGTGAGGAGTGGGCTTTGCCCAAGTGG 565
QY      1434 gaatccatatttggttgagactgaatcactcactgtgttactatcatctgtgtgtgtt 1493
        |||||||
Db      564 GAATCCTTNTTGGCTTACAGTATTCACACTTCCCTGTGATCACTTGTGGGTTTGTG 505
QY      1494 ttgttcgtttagaaaatttttgcgtgtgaanaacagcactccttggcttgaagcactg 1553
        |||||||
Db      504 TTGTTGCTGTAGAAATTTTTCGCTGTGAACAGSCAGCTCTTGTGGCTGGGCACTTG 445
QY      1554 tgtcattgcatgactgtgtgtgttccctcactcactccttctgatatgacaaataaagt 1613
        |||||||
Db      444 TGTCCATGCAATGAATCTGGGTGTTCCCTCCATCTCTTCTATTTGACAAATAATCAAGT 385
QY      1614 ttttcttcttcttgcacactcactgcatggcgtgaacacttcttcttcaacccctg 1673
        |||||||
Db      384 TGTTTTGTGTTTGTGTCMCTTACAGCATGGGCTAACCACTTCTTTTCAACCTCTG 325
QY      1674 aacacatttctctgattggttaactgacgaataattctatggaagaataacaggaagt 1733
        |||||||
Db      324 CACACCTTTTCTGTATGAGTGTGATGCGAGAAATATCTATTGGAAGAATCAGAGAACT 265
QY      1734 acaagtgcttcttgaacccctcctcaatggttcttagccttcaactctcacttcttct 1793
        |||||||
Db      264 ACCAATGCTTCTTGACCCCTTCTCAATGTTTCTAGCCCTTCACTCCATTTGCTTTTCT 205
QY      1794 gggcgtattcaagccctctgtgattcaactctgtctcactcactgtaactgtaacagag 1853
        |||||||
Db      204 -GGCTGATACAGAGCCCTGTGTGATCTTCCACACTCTGTGCTCCACTTTGATATCCGAG 146
QY      1854 tccaactgtaactgacagtgagctgctctctctggtgcaatgatacacaactgtaagtaact 1913
        |||||||
Db      145 ACCAACTTACTAATAACAGTACGTGCTTMTCTGAGACCATGATMAACACTGTAGAAGACT 86
QY      1914 aataactgcccagcctgggagagacaggaaggtcgtgatagtaagtaagttggttag 1973
        |||||||
Db      85 AATAACTGCCCAACCTGGGAGATCAGGAGATCTGCATGATTAAGTAAGTTGGTTTAC 26
QY      1974 ctttggtgtgcatcagtaactga 1998
        |||||||
Db      25 MTTTGTGTTGCAATCACTACTTA 1

```

**RESULT 4**  
**AL579274/c**

LOCUS AL579274 801 bp mRNA EST 16-FEB-2001  
 DEFINITION AL579274 LRT\_FL012.TC1 Homo sapiens cDNA clone CS0DH002YJ14 3 prime  
 / mRNA sequence.  
 ACCESSION AL579274  
 VERSION AL579274.1 GI:12944155  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 801)  
 Li, W. B., Gruber, C., Jesse, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91106 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
 FEATURES  
 source  
 1. 801  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CS0DH002YJ14"  
 /issue\_type="T cells from T cell leukemia"  
 /lab\_host="DH10B"  
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng liang life technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@life.com URL : http://fulllength.invitrogen.com 9 others

BASE COUNT 232 a 192 c 190 g 178 t  
 ORIGIN  
 Query Match 36.5%; Score 772.6; DB 10; Length 801;  
 Best Local Similarity 97.8%; Pred. No. 9; Se-124;  
 Matches 783; Conservative 9; Mismatches 8; Indels 1; Gaps 1;  
 1080 acaaggcgcctaagaacatcgagcgatgcatctgctgcagctgtaagatagca 1139  
 |||||||  
 Db 801 ACAAGGCCCTAATGAACATGAGCGGATGCGATGCGCATATGATGCA 742  
 |||||||  
 Qy 1140 gtgcagagaacaacagcactgtgcagaaagtgtgtgtactgttaagtgtgagag 1199  
 |||||||  
 Db 741 GTGCAGGAACAACAGCAGCTGCGAGCAAGTCTGTGTACTGTAAAGTGTGGAGG 682  
 |||||||  
 Qy 1200 cagagagagagcgagcgagcgctgagcttcacagatgagcaactgtggaacagagac 1259  
 |||||||  
 Db 661 CAGAGAGAGGAGGAGGCGCATGCGCTTCACAGCATGGCACATGTGGGACATGCGAGAC 622  
 |||||||  
 Qy 1260 attccctc-tcacagctagaactgaacaacacccctctgtcaggggtgtgcgtgtgaggt 1318  
 |||||||  
 Db 621 ATTCTCTTCCACAGCTAGATTAATGAACAACCCCTTCTGAGGGGTGTCGCTGAGGT 562  
 |||||||  
 Qy 1319 gtcatcctgtccccctcaataactaactagctggaactggcagcagccttactggct 1378  
 |||||||  
 Db 561 GTCATCTGTCGCCCTCATATTAATTAATGCTGGAATGCGACAGCCTCTACTGGGCT 502  
 |||||||  
 Qy 1379 ttactgtgagtgttgcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1438  
 |||||||  
 Db 501 TTYACTGTGATGTGTTCAGTTCATGCTCTAGGAAGTCAGCTTTTGGCCAGGAGGAATC 442  
 |||||||  
 Qy 1439 ctactgtgagctgaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1498  
 |||||||  
 Db 441 CTTATTTGGCTTAGCAGCTGATCCACTTCCATGTAACATCTGTGGGTTTGTGTTT 382  
 |||||||  
 Qy 1499 gctgttagaataatttgcgtgtgtgtaaaacagcagcagcagcagcagcagcagcagc 1558  
 |||||||

Db 381 GCCTGTAGAAATTTGTGGCTGTGTGAAGAAACAGCAGCTCTTTGGCTGAGACACTGTGCCC 322  
 Qy 1559 atgcagtactgtgtgttccctccatctcttcgatatgacccaataaagtgtt 1618  
 |||||||  
 Db 321 GTGCATGTACTGGGTGGTGGTCCCTCCATCCCTGATGATGACCAAAATCAAGTGTGTT 262  
 |||||||  
 Qy 1619 tgttttttccactcactgagcagtgaggaacacacttcttccaacccctcgaacac 1678  
 |||||||  
 Db 261 TGTGTTTTTTCACCTTCATGCGATGCGGCTAACCACTTTTTCATTAACCCCTGACAC 202  
 |||||||  
 Qy 1679 ctttttcagtgagtaacttcgaggaatacttcatttgaaagaagagagagag 1738  
 |||||||  
 Db 201 CTTTTCATGATGGGTAACTGACGAGGATATCTAATGGAAAGATACAGAGATACAG 142  
 |||||||  
 Qy 1739 tctcttccacccctcccaatgtttcagcctcactcactcactcactcactcactcact 1798  
 |||||||  
 Db 141 TGTCTTTTACCCCTTCTTCATATGTTTCTAGGCTTTCATCTTCATTTCTTTTCTGGGCT 82  
 |||||||  
 Qy 1799 gttactaacgctctgtgagatctcaactctgtcctccactcagtgatgagcagtcacca 1858  
 |||||||  
 Db 81 GATATACAGCCCTGTGTGATCTTCACACTCTGCTGCTCCATGATGAGCAGACACCA 22  
 |||||||  
 Qy 1859 cgtgaactgacagtgagtcgac 1879  
 |||||||  
 Db 21 CTGTAACTGACAGTGGCTGCC 1

RESULT 5  
 BI085329 863 bp mRNA EST 20-JUN-2001  
 LOCUS BI085329  
 DEFINITION 602870389F1 NIH\_MGC\_98 Homo sapiens cDNA clone IMAGE:5013406 5',  
 mRNA sequence.  
 ACCESSION BI085329  
 VERSION BI085329.1 GI:14503659  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 863)  
 NIH-MGC http://mhc.nhl.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLN at:  
 http://image.llnl.gov  
 Plate: LNCM1819 row: a column: 23  
 High quality sequence stop: 814.  
 Location/Qualifiers  
 1. 863  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5013406"  
 /clone\_lib="NIH\_MGC\_98"  
 /issue\_type="astrocytoma grade IV, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: brain; Vector: pOTB7, site: 1; XhoI; site: 2;  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCAGCAG(G). Library constructed by Ling Hong  
 in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."

BASE COUNT 212 a 226 c 192 g 233 t  
 ORIGIN



QY 481 tgaactgtttccaggaactactgatacgaagcattcttgagccccaacaacacatg 540  
 Db 483 TGTACCTGTTTCCAGGACACTAGCATGATCATTCTGTGAGCCCAAAA-AACAACCTG 541  
 QY 541 attctctagatactatcatatgcttccggaagcagctccaccagaattatagttatc 600  
 Db 542 ATTCTTTGATATATCATCTCTCCGAGAGAGTCCACCCAGAAATTTTGTATTT 601  
 QY 601 tagaaagatccctcctcattctctgagagactccggtgagcgtctgagacatctgt 660  
 Db 602 TAGAAAGGATATCCTCTCATTTCTGTGATGAGAGATCCGCTGCTGTGACATCTGT 661  
 QY 661 gacccaagatacagcgttgatcccaaccagaacatcatatcttgctctgagagagt 720  
 Db 662 GCACCAAGATACAGCGTGTGA-CCACCCAGCAATATGATATCTTGNGTCTGAGAGGT 720  
 QY 721 gttctctaacatctctcttgagatgaacca-actccagcgtttgca-cgtgaaagcct 778  
 Db 721 GTTCTCTAACCATCTCTGGGATGAACCAATCCAGGCTTTGCAACCGTGAAGCCTT 780  
 QY 779 gagccggacatgctctcacatcttaccctgctccctcctgttgagagacatggtgaa 838  
 Db 781 GAGCCGGGACATGCTTTCACATCTTACTGTCTCTCTCTCTGTTGAGAGACCTCGTTGAA 840

RESULT 7  
 LOCUS BG747535 819 bp mRNA EST 15-MAY-2001  
 DEFINITION 602704665F1 NIH\_MGC\_15 Homo sapiens cDNA clone IMAGE:4857880 5',  
 mRNA sequence.  
 ACCESSION BG747535  
 VERSION BG747535.1 GI:14058188  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapds-remail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: NIH Intramural Sequencing Center  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: L1CM1711 row: a column: 17  
 High quality sequence stop: 795.

FEATURES  
 source location/Qualifiers  
 1..819

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4857880"  
 /clone\_id="NIH\_MGC\_15"  
 /tissue\_type="adenocarcinoma cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: colon; Vector: pOT7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGACGAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)"  
 BASE COUNT 201 a 204 c 188 g 226 t

Query Match 34.8%; Score 736.6; DB 11; Length 819;

Best Local Similarity 97.2%; Pred. No. 1.5e-117;  
 Matches 792; Conservative 0; Mismatches 19; Indels 4; Gaps 4;

QY 1 ctgtgaagatgagcgtctctccaggtgtgtctggctgctgctgtgtgagctcgag 60  
 Db 4 CTGTGAAGATGGGCTCTCCA-6GTGTGCTGGGCTGGTCGGCTGTGTGGGCTCGGAG 62  
 QY 61 tcaaccctgacatcttgtaaccggaggtcgaacttgctgctctgctggcctgtgag 120  
 Db 63 TCACCCCTGGACATTTTGTACCCCGGAGGCTGCAACTTGTGCTCTGCTGCTGGG 122  
 QY 121 gggccctcgtgctctcaaaagcttcaaccttctccaaaggcaagtgtgaaagacttga 180  
 Db 123 GGCCCTCGCTGCTTCAAGCTTCACCTTTCTCCAAAGGAGATGTGAAGACTTGATGT 182  
 QY 181 ctatgtgttaaccaaagacaagaagcgttaagtggaataacatgcttctgtggtctc 240  
 Db 183 CTATGTGTGTAACCAAGCAAGCAAAAGCATTAATGGAAATACATCGTTTCTTGGGCTGC 242  
 QY 241 attcccccgtctctatctctgttacacatctctcaatgaaagatgtgagatgtatg 300  
 Db 243 ATTTCCCTCCCTCTTATGTCTGTACACATCTTCTATGAAGATTTGAGATTTATGSG 302  
 QY 301 ctgatacgaagcctagaagaataaagaacaatactgtgagaagacataataagtttc 360  
 Db 303 CTGATGCCAANAAGGCTAGAGATAAAGCAAAATGTGTGAAGGACATATTAAGTTTC 362  
 QY 361 atcaactcatabcggagagatgagacatttgagacagcttcgcgaagcgtccaccagt 420  
 Db 363 ATCAACTTCATACCGGAGATGAGACATTTGAGAGTTCCGCAAGACGTCCACCAAT 422  
 QY 421 gctcttccaggtatatttcatctcacccttctgcaactcactggtgctctgtccta 480  
 Db 423 GTCTTTTCCAGTATTTTTCATTTCCACTTTTCCCACTACTGCTGTCTTGTCTAA 482  
 QY 481 tgaactgtttccaggaactactgatacgaagcattcttgagccccaacaacacatg 540  
 Db 483 TGTACCTGTTTCCAGGACACTAGCATGATCATTCTGTGAGCCCAAAA-AACAACCTG 542  
 QY 541 attctctagatactatcatatgcttccggaagcagctccaccagaattatagttatc 600  
 Db 542 ATTCTTTGATATATCATCTCTCCGAGAGAGTCCACCCAGAAATTTTGTATTT 602  
 QY 601 ta-gaaagatccctcctcattctctgagagactccggtgagcgtctgagacatctg 659  
 Db 603 TAGGAAAGGATATCCTCTCATTTCTGTGATGAGAGATCCGCTGCTGTGACATCTG 662  
 QY 660 tgcacaagatacagcgttgatccc-accagaacatcatatcttgctctgagaga 718  
 Db 663 TGCACCAAGATACAGCGTGTGATCCCAACCCAGGAAATGATATCTTGCTGAGAGA 722  
 QY 719 gtgttctctaacatctccttgagagactggaacacacagcaggtttgcaagttaaagcct 778  
 Db 723 GTGTTCTCTAACCATCTCTGTGGCATGAACCA-TCAGGTTTGACGTTGAAAGCCTT 781  
 QY 779 gagccggacatgctctcacatcttaccctgctc 813  
 Db 782 GAGCCGGGACATGCTTTTCAAAATCTACCTGGCTC 816

RESULT 8  
 LOCUS BG746526 780 bp mRNA EST 15-MAY-2001  
 DEFINITION 602703825F1 NIH\_MGC\_15 Homo sapiens cDNA clone IMAGE:4857102 5',  
 mRNA sequence.  
 ACCESSION BG746526  
 VERSION BG746526.1 GI:14057179  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 780)









```

|||||
Db 301 CTTGGCTGTAATCTACGCATA-TGTTGAGATAGTGTGCAACTTGGCTGGAGAAATGG 359
Oy 996 ctgcagattcttcagcagctaaagaacgtacgtctctctctgtcgcacaagtgtc 1055
Db 360 CTGCAGATTCTCTGAGCCTGAAGAAGACCTGCTGCTCTCTCTGCTGCAACCTGCTC 419
Oy 1056 ctgtctccaccaactactcttggagacaagcgctgaatgaacatgagcgatgacat 1115
Db 420 CTGCTCTCCACCAACTACTTGGGACAGGGCGCTGAATGACACATGAGCGAGATGGCAT 479
Oy 1116 gtcccgagtgatagatagcagtcaggaacaacaacagcacttgcagacaagtctgt 1175
Db 480 GTCTCGAGCTGTATGATATGACACTGCAGAGAACAAACAGCACTTGGCCAAAGTCTGT 539
Oy 1176 gtctactgtaagtgtgtggagacagagagagagagagagagagagagagagagagat 1235
Db 540 GTGACTGTTAAGTGTGTGGAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 599
Oy 1236 ggcacacctgtggaaactgcagacatctctcacaagctgaactgaacaacacctctt 1295
Db 600 GGCACACATGTGGGAGACTGCAGACATTCCTCTCACAGCTGAACTGAACAAACCTCTT 659
Oy 1296 gctaggggtgtgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1338
Db 660 GCTAGGGGTGTGCTGCTGTGAGGTGTCTATCTCTCCCTCATATA 702

RESULT 14
LOCUS BG473855 714 bp mRNA EST 21-MAR-2001
DEFINITION 602515550F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4647143 5',
VERSION BG473855 mRNA sequence.
KEYWORDS BG473855.1 GI:13406132
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 714)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999).
COMMENT Contact: Robert Strausberg, Ph.D.
Email: sgabds@email.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLMH1423 row: d column: 24
High quality sequence sloop: 713.
Location/Qualifiers
1. 714
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4647143"
/clone_11b="NIH_MGC_16"
/tissue_type="retinoblastoma"
/label="host="DH10B (phage-resistant)"
/notes="Organ: eye; Vector: pOT7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCGAGG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."
BASE COUNT 165 a 166 c 191 g 192 t
ORIGIN

```

Query Match 32.4%; Score 686; DB 11; Length 714;  
 Best Local Similarity 98.5%; Pred. No. 8.2e-109;  
 Matches 703; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

```

Oy 890 gattggcagctgactgtctcaggaagaagtaaatgctgttctctcgtgtgctgaattc 949
Db 1 GATTGGCAGCTGACTGCTCAGGAAGTAATAATGGCGTTGTTACTCTCCGTGGCTGAATTC 60
Oy 950 taacgatattgtgaaatagtggtgtcgaacttggcttggagagaaatgtgctgaattccg 1009
Db 61 TACGCAATATTGTTGAAGATATAGTGTGCAACTTGGCTGGGAAATGGCTGAGATTTCTGT 120
Oy 1010 cagctgaaagaagctgtgagctgtctctctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1069
Db 121 CAGCTGAAAGAAGCTGAGCTGTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 180
Oy 1070 ctacacttggagacaagcgctgaaatgaacatggaatggaatggaatggaatggaatgga 1129
Db 181 CTACCTTTGGACAAAGGCGTGAATGAACCATGGAGGAGGATGGCATTTGCTGAGTCGTA 240
Oy 1130 taatatagcagtgcaagaaacaacagcacttgcagacaagctgtgtgtgtgtgtgtgtgt 1189
Db 241 TACTATAGCAGTGCAGAGAACAAACACACTTGGCCACAAAGTGTGTGTGTGTGTGTGT 300
Oy 1190 gtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1249
Db 301 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 360
Oy 1250 aactgcagacatctctctacagctgaactgaactgaactgaactgaactgaactgaactga 1309
Db 361 AACTGCAGACATTTCTCTCTACAGCTGAACCTGAACAAACCTCTGTGTGTGTGTGTGTGT 420
Oy 1310 gtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1369
Db 421 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480
Oy 1370 taactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1429
Db 481 TACTGTGGCTTTTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 540
Oy 1430 gtgggaatccttatttgcgttgaactgatacacttgcatacttgcatacttgcatact 1489
Db 541 GTGGGAATCCTTATTGTGGCTGTAGACTGATCCAAATTCATGTTACTTACATCTGTGGTC 600
Oy 1490 ttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1548
Db 601 ACAGTAGTAGCTGTGAAGAAATTTAGGCTGTGTGAAGACAGACACTCTTGGCTGGCAGC 660
Oy 1549 acttgttcatatcatcttactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1602
Db 661 ACTTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 714

RESULT 15
AL557670 702 bp mRNA EST 16-FEB-2001
LOCUS AL557670 LTI_FL012_Tc1 Homo sapiens cDNA clone CSDB007Y013 5 prime
DEFINITION , mRNA sequence.
ACCESSION AL557670
VERSION AL557670.1 GI:12901503
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 702)
AUTHORS Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France

```

Email: [segregenoscope.cns.fr](mailto:segregenoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr).  
Location/Qualifiers

## FEATURES

SOURCE

1. 702

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_id="LTI\_Fl012\_7C1"

/tissue\_type="T cells from T cell leukemia"

/lab host="DH10B"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : [fliang@lifetech.com](mailto:fliang@lifetech.com) URL : <http://fulllength.invitrogen.com>"

BASE COUNT 174 a 150 c 146 g 231 t 1 others  
ORIGIN

Query Match 32.2% Score 681; DB 10; Length 702;  
Best Local Similarity 98.9% Pred. No 6e-108;  
Matches 695: Conservative 1; Mismatches 6; Indels 1; Gaps 1;

```
OY 1397 gttcatgtctctaggaagtcagcgtttgtcccaagtgaggatccttatttgcttaggaact 1456
    |||||||
Db 1 gtttcattgtcttaggaagtcagcgtttgtcccaagtgaggatccttatttgcttaggaact 60

OY 1457 gatcaactccatgttacttaacatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1516
    |||||||
Db 61 gatccactttccatgttacttaacatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 120

OY 1517 gctgtgtgaaacagcagcctcttgctgagagacttgttccatgacatgtacttgggtgt 1576
    |||||||
Db 121 gctgtgtgaaacagcagcctcttgctgagagacttgttccatgacatgtacttgggtgt 180

OY 1577 ttccctccatccttcttgatgatgaacaaatcaagtgtgtgtgtgtgtgtgtgtgtgt 1636
    |||||||
Db 181 ttccctccatccttcttgatgatgaacaaatcaagtgtgtgtgtgtgtgtgtgtgtgt 240

OY 1637 ctggagtgagcttaacacactcttttcaaacctctgaacacactcttctctatgagtgtaac 1696
    |||||||
Db 241 ctggagtgagcttaacacactcttttcaaacctctgaacacactcttctctatgagtgtaac 300

OY 1697 ttgcaggaatattctatltggaagaagaatacaggaagtacagtgctctcttgaccccttcc 1756
    |||||||
Db 301 ttgcaggaatattctatltggaagaagaatacaggaagtacagtgctctcttgaccccttcc 360

OY 1757 tcaatgttcttagccttcaactctcaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1816
    |||||||
Db 361 tcaatgttcttagccttcaactctcaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 420

OY 1817 gatcttcaactctgctgacctccactgtgacagcagtcacacatgtaactgacagtgagct 1876
    |||||||
Db 421 gatcttcaactctgctgacctccactgtgacagcagtcacacatgtaactgacagtgagct 480

OY 1877 gaccttcttgagcagatgacacactgttaaggtactaattactgcccagccttgaggaga 1936
    |||||||
Db 481 gaccttcttgagcagatgacacactgttaaggtactaattactgcccagccttgaggaga 539

OY 1937 tcaggaagagtgctgcatagtgtagtaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1996
    |||||||
Db 540 tcaggaagagtgctgcatagtgtagtaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 599

OY 1997 tagagttctgtaataacttattgaaatgcatgaagcacatgtttttaaaccacagtaag 2056
    |||||||
Db 600 tagagttctgtaataacttattgaaatgcatgaagcacatgtttttaaaccacagtaag 659

OY 2057 actgcttgaaacctgttgatgagtaaaaaaa 2099
    |||||||
Db 660 actgcttgaaacctgttgatgagtaaaaaaa 702
```

Search completed: December 20, 2001, 17:58:31  
Job time: 6568 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2001, 17:25:04 ; Search time 99.76 Seconds

(without alignments)  
4808.344 Million cell updates/sec

Title: US-09-868-474-1

Perfect score: 2118

Sequence: 1 ctgtgaagatgagcgtctcc.....aaaaaaaaaaaaaaaaaaaaa 2118

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Maximum Match 1008  
Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/1na/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/1na/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/1na/PCRTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/1na/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	2.6	7218	1 US-08-232-463-14	Sequence 14, Appl
2	54.6	2.6	3138	1 US-07-867-106-4	Sequence 10, Appl
3	51.2	2.4	1582	3 US-08-545-196B-10	Sequence 12, Appl
4	51.2	2.4	1582	3 US-08-545-196B-12	Sequence 12, Appl
5	51	2.4	2671	6 5168051-9	Patent No. 5168051
6	50.2	2.4	2744	3 US-09-071-101-1	Sequence 1, Appl
7	50.2	2.4	2744	3 US-09-369-618-1	Sequence 1, Appl
8	50.2	2.4	2744	3 US-09-369-617-1	Sequence 1, Appl
9	49.8	2.4	1522	4 US-09-413-574-1	Sequence 1, Appl
10	49.8	2.4	3736	2 US-08-480-473B-1	Sequence 1, Appl
11	49.8	2.4	3736	3 US-08-915-213-1	Sequence 1, Appl
12	49.8	2.4	3736	3 US-09-148-547-1	Sequence 1, Appl
13	49.8	2.4	3736	4 US-09-235-217-1	Sequence 1, Appl
14	49.8	2.4	3736	5 US-09-966-10251-1	Sequence 1, Appl
15	49.6	2.3	2296	4 US-08-946-841C-137	Sequence 13, Appl
16	49.4	2.3	375	3 US-08-946-026-23	Sequence 23, Appl
17	49.4	2.3	2567	3 US-08-995-260-4	Sequence 23, Appl
18	49.4	2.3	2628	1 US-08-143-219-1	Sequence 1, Appl
19	49.4	2.3	3581	2 US-08-738-349-1	Sequence 1, Appl
20	49.2	2.3	1540	3 US-08-977-001-2	Sequence 2, Appl
21	49.2	2.3	2852	2 US-09-027-137-2	Sequence 2, Appl
22	49	2.3	3527	2 US-08-909-965C-7	Sequence 7, Appl
23	49	2.3	3637	1 US-08-445-640-3	Sequence 3, Appl
24	49	2.3	3637	1 US-08-170-558-3	Sequence 3, Appl
25	49	2.3	3637	3 US-08-447-314-3	Sequence 3, Appl
26	49	2.3	3637	3 US-08-445-661-3	Sequence 3, Appl
27	48.8	2.3	966	1 US-08-514-014-7	Sequence 7, Appl

#### ALIGNMENTS

28	48.8	2.3	966	2 US-08-833-823-7	Sequence 7, Appl
29	48.8	2.3	1393	1 US-08-174-467-18	Sequence 18, Appl
30	48.8	2.3	1393	3 US-08-452-071-18	Sequence 18, Appl
31	48.8	2.3	1804	2 US-08-504-459-5	Sequence 5, Appl
32	48.4	2.3	1315	4 US-09-721-822A-10	Sequence 10, Appl
33	48.2	2.3	467	2 US-08-841-349-18	Sequence 18, Appl
34	48.2	2.3	1931	3 US-09-019-942-2	Sequence 2, Appl
35	48.2	2.3	2255	1 US-08-871-572B-3	Sequence 3, Appl
36	48	2.3	208	1 US-08-686-878A-37	Sequence 37, Appl
37	48	2.3	1772	2 US-08-960-022-13	Sequence 13, Appl
38	48	2.3	2301	4 US-09-232-191-8	Sequence 8, Appl
39	48	2.3	2301	4 US-09-232-200-8	Sequence 8, Appl
40	48	2.3	2301	4 US-09-232-197-8	Sequence 8, Appl
41	48	2.3	2710	4 US-09-232-200-44	Sequence 44, Appl
42	48	2.3	2710	4 US-09-232-200-70	Sequence 44, Appl
43	48	2.3	2710	4 US-09-232-197-44	Sequence 44, Appl
44	48	2.3	2710	4 US-09-232-197-70	Sequence 44, Appl
45	47.8	2.3	3785	1 US-08-485-618-98	Sequence 98, Appl

RESULT 1  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEFFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOMLOPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)836-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: PTZgpt-Fls  
US-08-232-463-14



Db 1536 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1567

RESULT 5  
5168051-9

```

1  APPLICANT: DERYNCK, RIK M.A.; GOEDEL, DAVID V.
2  TITLE OF INVENTION: NUCLEIC ACID ENCODING TGF-B ITS USES
3  NUMBER OF SEQUENCES: 21
4  CURRENT APPLICATION DATA:

```

```

      APPLICATION NUMBER: 05/07/389,929
      FILING DATE: 04-AUG-1989
      SEQ ID NO:9:
      LENGTH: 2671

```

Query Match 24%. Score 51. nb 6. result 20

Matches	84;	Conservative	0;	Mismatches	55;	Indels	
Oy	1980	ttgtgtcatcagtgacttaagttctgtaataa	cttattatataatgcata				

Db	2488	tcttctgtataataataataatttttggggaatccttgatcatattcatcttctt
Oy	2040	tttaaacccaagtaagaactgccttgaacgcgttgatgtaaaaaaaaaaaaaa

```
Db      2608 aaaaaaaaaaaaaaaaaa 2626
```

US-09-071-101-1  
; Sequence 1, Application US/09071101  
; Patent No. 6013503  
GENERAL INFORMATION.

APPLICANT: LOK, SI  
APPLICANT: Jaspers, Stephen R.  
TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4  
NUMBER OF SEQUENCES: 11

ADDRESSEE: Zymogenetics, Inc.  
STREET: 1201 Eastlake Avenue East  
CITY: Seattle

```

;
;      COUNTRY:  USA
;      ZIP:      98102
;
COMPUTER READABLE FORM:

```

COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0

APPLICATION NUMBER: US/09/071,101  
FILING DATE:  
CLASSIFICATION:  
PRIORITY DATE:

APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Barber, Cary D

REGISTRATION NUMBER: 31,648  
REFERENCE/DOCKET NUMBER: 97-05  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 205-442-6673

TELEX: 200-442-00/8  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:

QY 2094 aaaaaaa 2118





```

ADDRESSSEE: Darby & Darby, PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/496,841C
FILING DATE: 28-Jun-1995
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Paul F. Fehlner, Ph.D.
REGISTRATION NUMBER: 35,135
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 527-7700
TELEFAX: (212) 753-6237
INFORMATION FOR SEQ ID NO: 137:
SEQUENCE CHARACTERISTICS:
LENGTH: 2296 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 137:
US-08-496-841C-137

```

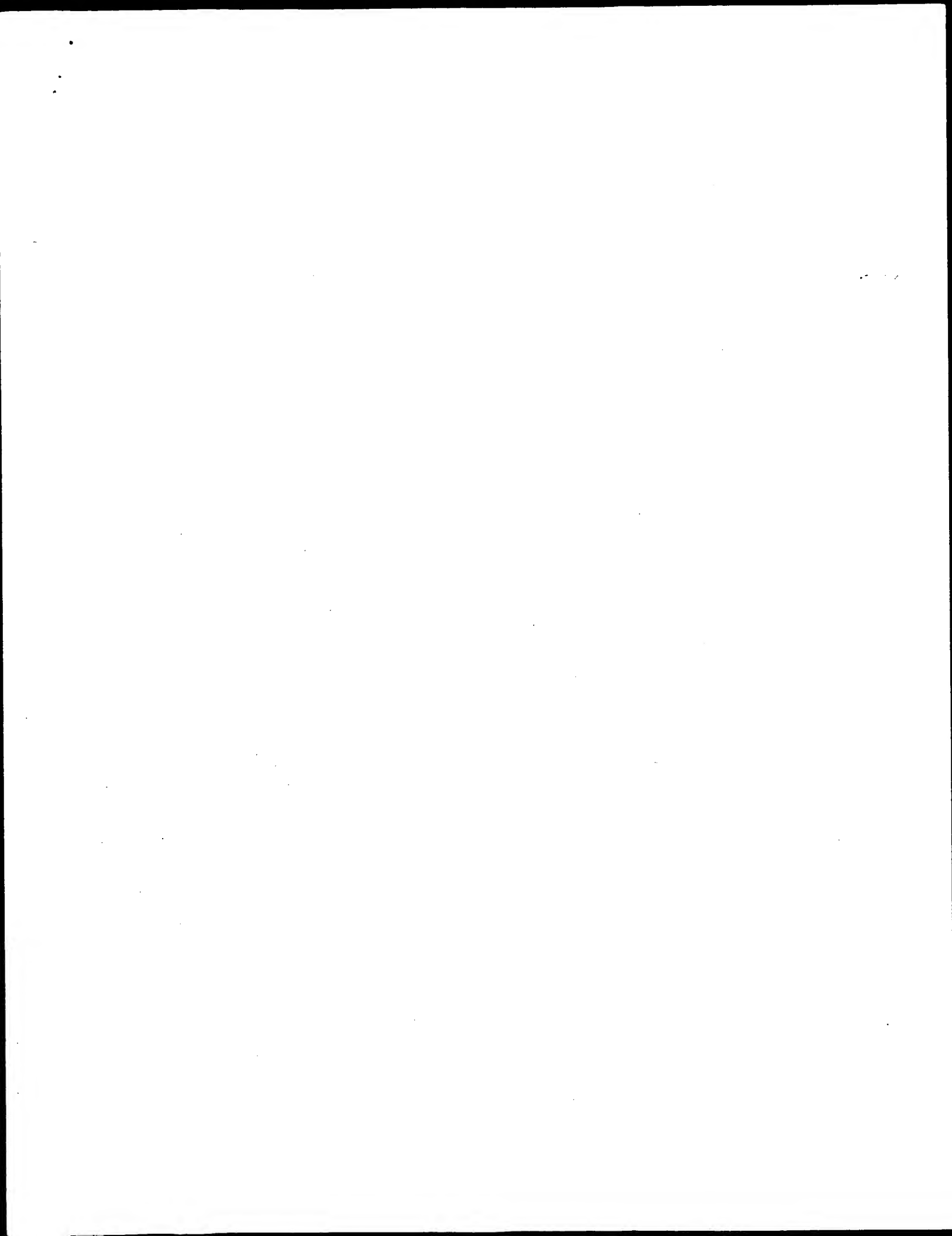
```

Query Match      2.3%; Score 49.6; DB 4; Length 2296;
Best Local Similarity 68.8%; Pred. No. 0.0002;
Matches 55; Conservative 4; Mismatches 21; Indels 0; Gaps 0;

QY 2039 ttltaaaccaagtaagactgcttgaacctgtgataggaaaaaaaaaaaaaaaaaaa 2098
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 2198 TMANARTNANANCCAAWAAATANNNTNANTCTRTTANNAAAAAAAAAAAAAAAAAAA 2257
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2099 aaaaaaaaaaaaaaaaaaa 2118
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2258 AAAAAAAAAAAAAAAAAAA 2277

```

Search completed: December 20, 2001, 18:51:31  
 Job time: 5187 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2001, 10:09:52 ; Search time 15.97 Seconds

(without alignments)  
1717.147 Million cell updates/sec

Title: US-09-868-474-2

Perfect score: 1912

Sequence: 1 MALSRVCMARSAMGSAVTP.....LSLLHNVLSTNYLGR 360

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR:\*\*\*  
2: PIR2:\*\*\*  
3: PIR3:\*\*\*  
4: PIR4:\*\*\*

Prog. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	520	27.2	100	2	T08763	hypothetical prote
2	150.5	7.9	398	2	C96680	hypothetical prote
3	143	7.5	485	2	T38255	hypothetical prote
4	140	7.3	770	2	T22944	hypothetical prote
5	140	7.3	784	2	T22939	hypothetical prote
6	136.5	7.1	755	2	T47806	hypothetical prote
7	136	7.1	753	2	S66710	probable membrane
8	131.5	6.9	497	2	T21579	hypothetical prote
9	131	6.9	481	2	T42226	hypothetical prote
10	119.5	6.2	454	2	S69017	probable membrane
11	101.5	5.3	431	1	C69689	response regulator
12	99	5.2	3795	2	T00831	hypothetical prote
13	95	5.0	369	2	B69799	hypothetical prote
14	92.5	4.8	728	2	A81385	salicylate 1-mono
15	92	4.8	656	2	H71867	probable ATP /GTP
16	91.5	4.8	1120	2	S46111	methionine--tRNA
17	90	4.7	880	2	S61908	probable membrane
18	89	4.7	276	2	S35270	DNA-binding protei
19	88.5	4.6	287	2	J01743	DNA-damage repair
20	88	4.6	1163	2	T24855	hypothetical prote
21	87	4.6	314	2	G69818	hypothetical prote
22	87	4.6	735	2	D96737	CMF-binding factor
23	86.5	4.5	254	2	T43494	ARE1-like protein,
24	86.5	4.5	957	2	C64463	hypothetical prote
25	86	4.5	1715	2	T06145	type I restriction
26	85.5	4.5	679	2	T20713	disease resistance
27	85.5	4.5	2231	2	S53416	hypothetical prote
28	84	4.4	561	2	F86561	SEMI protein - yea
29	84	4.4	561	2	C72063	arginyl tRNA trans
						arginyl tRNA synth

30	84	4.4	650	2	A64572	methionine--tRNA
31	84	4.4	701	2	G96737	ARE1-like protein,
32	84	4.4	2165	2	T21371	hypothetical prote
33	83.5	4.4	339	2	T15251	hypothetical prote
34	83.5	4.4	465	2	T48374	UDP-glucosyltrans
35	83.5	4.4	603	2	B86608	exonuclease ABC,
36	83.5	4.4	603	2	F72017	hypothetical prote
37	83.5	4.4	772	2	T16474	hypothetical prote
38	83	4.3	1433	2	A71444	probable LTR retro
39	82.5	4.3	2199	2	T40008	Cdc20P - fission y
40	82	4.3	248	2	H83111	hypothetical prote
41	82	4.3	1596	2	T24146	hypothetical prote
42	81.5	4.3	816	2	T25555	hypothetical prote
43	81.5	4.3	884	2	F96659	protein F2k1.26 f
44	81.5	4.3	1511	2	T42711	sulfonylurea recep
45	81.5	4.3	1546	2	T42728	sulfonylurea recep

#### ALIGNMENTS

RESULT 1  
T08763  
hypothetical protein DKFZp586A011.1 - human  
C:Species: Homo sapiens (man)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 04-Mar-2000  
C:Accession: T08763  
R:Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wilmann, S.  
Submitted to the Protein Sequence Database, May 1999  
A:Reference number: 216468  
A:Accession: T08763  
A:Molecule type: mRNA  
A:Residues: 1-100 <KOE>  
A:Cross-references: EMBL:AL050286  
A:Experimental source: adult uterus; clone DKFZp586A011  
C:Genetics:  
A:Note: DKFZp586A011.1  
C:Superfamily: human hypothetical protein DKFZp586A011.1

Query Match 27.2%; Score 520; DB 2; Length 100;  
Best Local Similarity 100.0%; Pred. No. 3.7e-38;  
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 MLTSTYPPPLRLRHUKHTTYTHOLDKALAKLGIGQLTAQEVKSACYLRGINSTHIGED 320  
DB 1 MLTSTYPPPLRLRHUKHTTYTHOLDKALAKLGIGQLTAQEVKSACYLRGINSTHIGED 60  
QY 321 RCRTWLGEWLQTSCKSEAEKSLHLHNVLSTNYLGR 360  
DB 61 RCRTWLGEWLQTSCKSEAEKSLHLHNVLSTNYLGR 100

RESULT 2  
C96680  
hypothetical protein F5114.7 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: C96680  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federzpiel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
Nansen, N.F.; Hughes, B.; Hultzer, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, T.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marzia,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shin, P.; Southwick, A.M.; Sun, H.; Tallo,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: C96680  
A>Status: preliminary  
A:Molecule type: DNA



A:Introns: 44/2; 184/3; 360/2; 708/3; 757/3

## Query Match

Best Local Similarity 7.3%; Score 140; DB 2; Length 784;  
Matches 57; Conservative 52; Mismatches 113; Indels 24; Gaps 7;

QY 120 LPYREMEHLROFRODVTKCLFLGITSIPPANVLPFLMLTSPFROLIRHFKTHHTVYHQL 179  
DB 129 LSRERROOLVATVSDFLVLPFSEFIIIVPMELALPFIPLFGOMLPSPFOESSKEE-- 185  
QY 180 LDIYHARPKQ--SHPEIISYLEKVIPLI-----SDAGLRWRLTDLCTYIQ-RGTH 226  
DB 186 ----EKRRKQVKLVREMAKFLDPIEIGLERKTRNKESTRSLFAL--FIKKVNEGCV 239  
QY 227 PAINDILRECFSNHPLGKNOLOALHVKALSRAMLLTSTLPPLIRHKLKTHHTVYHQL 286  
DB 240 VSNELLKFKSLPEDE-ITLDNLSMGOLRSLCKMLKSTNSGSEIILRQNLNKRIRLQAD 298  
QY 287 DKAL-AKLGIGOLTAOEKVSACVLRGLNSTHIGEDRCRTWLGEMLOISLKEAELSLIL 345  
DB 299 DKQIAEGGVADLSSIDLQSAACRARGMRAIGVSEERLQOLVQWLELINDKVPALLIL 358  
QY 346 HNVVLL 351  
DB 359 SRTLYL 364

## RESULT 6

T47806  
Hypothetical protein F24G16.90 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C:Accession: T47806  
R:D'Angelo, M.; Verzé, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Lemcke, K.;  
submitted to the Protein Sequence Database, February 2000  
A:Reference number: 224477  
A:Accession: T47806  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-755 <DAN>  
A:Cross-references: EMBL:AL138647  
A:Experimental source: cultivar Columbia; BAC clone F24G16  
A:Map position: 3  
A:Introns: 194/2; 298/3; 347/3; 377/2; 408/2; 450/3; 472/2; 530/3; 595/3; 611/3; 660/3;  
A>Note: F24G16.90

## Query Match

Best Local Similarity 7.1%; Score 136.5; DB 2; Length 755;  
Matches 50; Conservative 56; Mismatches 123; Indels 25; Gaps 6;

QY 92 GLQMLWADAKKARRITKNNKHNKPHQLPYREMEHLROFRODVTKCLFLGITSIPPAN 151  
DB 218 GTKLIMADTRISSRLKL-----AGKSLSRERROOLVATVSDFLVLPFSEFIIIVPMEL 273  
QY 152 YIVFELMLYLPFROLIRHFKTHHTVYHQLKALAKLIGOLTAOEKVSACVLRGLNSTHIG 209  
DB 274 FLVLPFSEFIIIVPMELALPFIPLFGOMLPSPFOESSKEE--SDAGLRWRLTDLCTYIQ-RGTH 326  
QY 210 LNRW-----LTDLCTKIORGTHPAIHDLALRECFSNHPLGKNOLOALHVKALSR 259  
DB 327 KHSRTGEVQQTAEDEDFLDKVRKGOIVNDELDFPAKLF-NDELTLNISPRLVSMCK 385  
QY 260 AMLLTSTYLPPLIRHKLKTHHTVYHQLKALAKLIGOLTAOEKVSACVLRGLNSTHIG 319  
DB 386 YKISIPYGDVAVLRWLRKRLRSIKEDDKLIRAGVDSLSAEELREDCREGMLGL-VSV 444  
QY 320 DRCRTWLGEMLOIS 333  
DB 445 EEWROQLRDMMDLS 458

## RESULT 7

S66710  
Probable membrane protein YOL027c - Yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein O2165  
C:Species: Saccharomyces cerevisiae  
C>Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 05-Nov-1999  
C:Accession: S66710  
R:Habbig, B.; Hattenhorst, U.; Hollenberg, C.P.; Ramezani Rad, M.  
submitted to the Protein Sequence Database, July 1996  
A:Reference number: S66703  
A:Accession: S66710  
A:Molecule type: DNA  
A:Residues: 1-573 <HAB>  
A:Cross-references: EMBL:Z74769; NID:g1419812; PID:e251846; PID:g1419813; GSPDB:GN000  
A:Experimental source: Strain S288C  
C:Genetics:  
A:Gene: MIPS:YOL027c  
A:Map position: 15L  
C:Keywords: transmembrane protein  
F:139-155/Domain: transmembrane #status predicted <TM>

Query Match  
Best Local Similarity 7.1%; Score 136; DB 2; Length 573;  
Matches 68; Conservative 58; Mismatches 117; Indels 50; Gaps 12;

QY 89 FMKGLQMLWADAKKARRITKNNKHNKPHQLPYREMEHLROFRODVTKCLFLGITS 145  
DB 94 YANGTLLGVEIKVSTKL-----IKFAQGYLSRERROOLVATVSDFLVLPFSEFIIIVPMEL 146  
QY 146 IPPPANYLVFLMLYLP-----ROLIRHFKTHHTVYHQLKALAKLIGOLTAOEK 189  
DB 147 IIPFELFLPFAKLFEPNLDPSTYESGMDQAKRKLIF---IKKTSSEF---HETLEE 200  
QY 190 SHPEIISYLEKVIPLISDAGLRWRLTDLCTYIQRGTHPAI---HD-IPLARECFSNHPL 244  
DB 201 SN-LITV-----NTIENAEKKOKFLNFPRKLYSAKEGKIWFQHDHISAIQMFKNDSV 253  
QY 245 GNOLOALHVKALSRAMLLTSTLPPLIRHKLKTHHTVYHQLKALAKLIGOLTAOEK 304  
DB 254 -LDNLSRPOLAAMSKFMSLRFGNDNMIRQIRKLDIMNDKDTIDYEGVESLSQEEY 312  
QY 305 SACYLRGKINSTHIGEDRCRTWLGEMLOISLKEAELSLHNVVLLSTNYLG 357  
DB 313 QACVSRKKAIGVSKEDLVNLIKWLDEL--RLRQKIPSVLM--VLSPTTFEG 360

## RESULT 8

T21579  
Hypothetical protein F30F8.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000  
C:Accession: T21579  
R:Milkinson, J.  
submitted to the EMBL Data Library, October 1996  
A:Reference number: Z19443  
A:Accession: T21579  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-497 <MIT>  
A:Cross-references: EMBL:Z81073; PID:CA803034.1; GSPDB:GN00019; CESP:F30F8.3  
A:Experimental source: clone F30F8  
C:Genetics:  
A:Gene: CESP:F30F8.3  
A:Map position: 1  
A:Introns: 36/3; 109/1; 189/1; 235/2; 276/1; 314/3; 342/2; 384/3; 404/3; 451/3; 479/1  
C:Superfamily: Caenorhabditis elegans hypothetical protein F30F8.3

Query Match  
Best Local Similarity 6.9%; Score 131.5; DB 2; Length 497;  
Matches 62; Conservative 51; Mismatches 127; Indels 55; Gaps 8;

RESULT 9  
T42226  
Hypothetical protein - f1eslon yeast (Schizosaccharomyces pombe) (fragment)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jul-2000  
C:Accession: T42226  
R:Yoshiohka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.  
DNA Res. 4, 363-369, 1997  
A:Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs  
A:Reference number: Z1323; M01D:98162722  
A:Accession: T42226  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-481 <EOS>  
A:Cross-references: NID:g1749441; PIDN:BAAL3779.1; PTD:g1749442  
A:Experimental source: strain PR745

C:Date: 22-Aug-1996 #sequence\_revision 06-Sep-1996 #text\_change 23-Mar-2001  
C:Accession: S69017  
R:Johnson, D.  
submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of *S. cerevisiae* cosmid 9642.  
A:Reference number: S69014  
A:Accession: S69017  
A:Molecule type: DNA  
A:Residues: 1-54 <JOH>  
A:Cross-references: EMBL:U40828; NID:q1066467; PID:q1066469; GSPDB:GN00016; MIPS:YPR1  
C:Genetics:  
A:Gene: MIPS:YPR125w  
A:Map position: 16R  
C:Keywords: transmembrane protein  
F:137-153/Domain: transmembrane #status predicted <TMM>

RESULT 11  
C69689  
response regulator aspartate phosphatase raph - Bacillus subtilis  
C:Species: *Bacillus subtilis*  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
C:Accession: C69689; 140009  
R:Kunst, F.; Ogatawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
C.; Bron, S.; Brönnliet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gal  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gal  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau  
y, M.; Ogawa, K.; Ogilwara, A.; Oudega, A.; Oudega, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scall  
Rieger, M.; Rivolta, C.; Rocha, E.; Rochoe, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se  
A:Authors: Scheich, S.; Schroeter, R.; Scroffone, F.; Sekiguchi, J.; Sekowska, A.; Se  
akuchii, M.; Tanakoshi, A.; Tanaka, T.; Terpetra, P.; Tomonoi, A.; Tosato, V.; Uchiya  
T.; Winters, P.; Wipart, A.; Yamamoto, H.; Yamane, K.; Yasumori, K.; Yata, K.; Yoshida  
A:Authors: Yoshikawa, H.F.; Zumsel, E.; Yoshikawa, H.; Yashimori, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*  
A:Reference number: A69580; MUID:198044033  
A:Accession: C69689

A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-3795 <STO>  
A:Cross-References: GB:AE002093; NID:g6598526; PIDN:AAD2014.2; GSPTB:GN00139  
C:Genetics:  
A:Gene: At2g17930; T13L16.5  
A:Map position: 2  
A:Protons: 20/1; 258/3; 287/2; 323/3; 355/3; 376/3; 396/3; 419/3; 441/1; 458/3; 467/3; 8

202 -----IP LISDAGLRWRLTDCTKIQGTHTPAIHDI LALRECF SNHPLGMNQLQ 251

Db 270 AVOEPWKKGVIIIGDAHAGAPF-----LAOGAAMAEIAIYLAELQNHADHETALQA 324  
 QY 252 LHVKALSSAMLLTGYLPPLRLRHTKTHHTVYHQDLAKLAKLGIGOLTAQEVKSACY 308  
 Db 325 YKRRAPRALKVON-LSEIVRRRLK-----GEGAEELIGECY 362

## RESULT 14

A81385  
 Probable ATP /GTP binding protein Cj0411 [Imported] - Campylobacter jejuni (strain NCTC C:Species: Campylobacter jejuni  
 C:Date: 31-Mar-2000 #sequence.revision 31-Mar-2000 #text\_change 31-Mar-2000  
 C:Accession: A81385  
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chillingworth, C.W.; O'Neill, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barré-Naulte 403, 665-668, 2000  
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyf  
 A:Reference number: A81250; MUID:20150912  
 A:Accession: A81385  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-728 <PAR>  
 A:Cross-references: GB:AL139075; GB:AL11168; NID:96967817; PIDN:CAB74247.1; PID:9696788  
 A:Experimental source: serotype O2, strain NCTC 11168  
 C:Genetics:  
 A:Gene: Cj0411

Query Match 4.8%; Score 92.5; DB 2; Length 728;  
 Best Local Similarity 22.1%; Pred. No. 5;

Matches 79; Conservative 47; Mismatches 115; Indels 117; Gaps 17;

QY 82 FYLLYTFKGLQMLMADAKKARIKTMKKHNKTFHQLPYREMEHLRQFRODYTKCL-- 139  
 Db 128 FKELNTINLE-LQKLSFD-----ENLKSRLQKLEKQNL-----EFNIAITGVMA 173  
 QY 140 -----FLGISIRPFANYVFLMYLFPROLIRHFWTPKQGFDFDIYHAF 186  
 Db 174 GKSSLLMALKEDELGVSNIPETANLV--LSYKSEAKI-YFMDKKEKMONILESSH- 229  
 QY 187 RKQSHPEIISYLEKY-----PLISDACL-----RWRLDLCRTIORGTH-- 226  
 Db 230 -NADLKEFLDKDSVNIEDFIKDKPLIONIALCELNFSKAKNISALIKKIEIKSHLE 288  
 QY 227 -----PAIDITLALRECFSNHPLG-----MNOQLAHVAKLSRAMLLISYL 267  
 Db 289 FLKNNISIVDPGLDGVVQREIVTNEYLRSDFLIHLNMAOSLSLQKD-----ADFL 341  
 QY 268 PPELRLRRLKTHHTVYHOLDKALAKLGIGOL--TAQEVKS----- 305  
 Db 342 VHCILNSRLSKFLIVLTKAD-LLSKKDLEVIYVTKESKSLVLDENLVEKIDFLCVS 400  
 QY 306 ----ACYLGLNSTH-----IGEDRCRTWL-----GEMLIQCSLKEAFLSLHLNVY 349  
 Db 401 AKMASDEYKGLASKESLOKSGMQEPENYLFENELVAGKSKIALRAYKKELHLELNIL 458

## RESULT 15

H71867  
 methionine--tRNA ligase (EC 6.1.1.10) - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori  
 A:Variety: strain J99

C:Date: 12-Feb-1999 #sequence.revision 12-Feb-1999 #text\_change 16-Jul-1999

C:Accession: H71867  
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;

Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;  
 Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path

A:Reference number: A71800; MUID:99120557  
 A:Accession: H71867  
 A:Status: preliminary  
 A:Molecule type: DNA

A:Residues: 1-656 <ARN>  
 A:Cross-references: GB:AE001525; GB:AE001439; NID:g4155533; PIDN:AAD06543.1; PID:g415  
 A:Experimental source: strain J99  
 C:Genetics:  
 A:Gene: metG  
 C:Superfamily: methionine--tRNA ligase  
 C:Keywords: aminocacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 4.8%; Score 92; DB 2; Length 656;  
 Best Local Similarity 17.7%; Pred. No. 4.8; Indels 64; Gaps 8;

Matches 49; Conservative 51; Mismatches 113;

QY 28 LQGRSLAMGAPRSSKLHLSPKADVKKMSYVYTKKAINGYHFRFLGRHPPFYLYT 87  
 Db 204 LSTITSEKGLPLPKKN-DPKHVVYVWDALNLNVSALG----- 243  
 QY 88 IFMKGLQMLMADAKKARIKTMKKHNKTFHQLPYREMEHLRQFRODYTKCLFLGIISIP 147  
 Db 244 -YLNGLDNKMAHFERARH-----VGKDLIRFAIYWP-----AFLMSLNP 284  
 QY 148 PRANYLVFLMYLFPROLIRHFWTP-----PKQGFDFDIYHAFRKQSHPEIISYLEKY 201  
 Db 285 LF-----KQLCVGWMTIEGVKSKSLGVNLAQKIAMRYGIEELRYFLRE 331  
 QY 202 IPLISDAGLRWR--LTDLCRTIORGTHPAIDILALRECFSNHPLGNNOLALHVAKLSR 259  
 Db 332 VFGGDGDSKKAIVERINANLNDLGNLRLGMAKKYFNYSLKSTKITAVYKPELEK 391  
 QY 260 AMLL-----TSYLPPLRLRRLKTHHTVYHOLDKALAK 292  
 Db 392 AHQILDNANSFVPMQQLKALEELFNIDYDLNKLIAK 428

Search completed: December 20, 2001, 10:12:00  
 Job time: 128 sec